GUSTAVE L.
AND
JANET W.
LEVY

GUSTAVE L.
AND
JANET W.
LEVY
LIBRARY

Digitized by the Internet Archive
in 2015

https://archive.org/details/mountsinaijourna7526moun

MOUNT SINAI JOURNAL OF MEDICINE

VOL. 75, NO. 2
MARCH/APRIL 2008

A Journal of Translational and Personalized Medicine

Issue Theme
9/11 World Trade Center Disaster, Past and Future

ntroduction by John Howard, MD,
Director, National Institute for
Occupational Safety and Health

WTC Medical Monitoring and
Treatment Program:
A Comprehensive Response

Epidemiologic Research on
Man-Made Disasters:
Strategies and Implications

Upper and Lower Respiratory
Diseases after WTC-like Disasters

Standardization of Spirometry in
Post-disaster Assessments

Diffuse Parenchymal Lung Disease as
Potential Sequel to WTC Exposures

Post-disaster Mental Health of
9/11 Workers and Volunteers

Impacts of WTC Disaster on
Children and Pregnant Women

Responder Safety and Health:
Preparing for Future Disasters

Special Clinical Features

Making Sense of ENHANCE: How Aggressively Should We Lower LDL Cholesterol?
Applying Evidence-based Medicine to Current Clinical Problems
Neurotherapeutics in Multiple Sclerosis

WILEY-BLACKWELL

Discover
Something
Great

Distinctive Content

Content
Journals • Backfile Collections
Reference Works & Encyclopedias
OnlineBooks™
Current Protocols Lab Manuals
Evidence-Based Resources

Access
Single article/chapter or full title
One-time fee, subscription,
or Pay-Per-View options
Unlimited concurrent usage

Capabilities
E-mail Alerts • EarlyView®
Citation download & tracking
Remote access • Mobile Edition®

Promote the process of scientific discovery with a different breed of research system: **Wiley InterScience**®. Long distinguished by its scope and extensive coverage of emerging disciplines, Wiley's authoritative content spans the spectrum of scientific, medical, technical, and business specialties. The dynamic **Wiley InterScience** platform adds value to your research by providing simple desktop access to this preeminent digital library of journals, reference works, books, laboratory manuals, evidence-based databases, and more. Researchers reap the combined benefits of top-tier resources and up-to-the-minute functionality. Experience the difference **Wiley InterScience** can make at your institution today.

WILEY
InterScience®
DISCOVER SOMETHING GREAT

WILEY
1807-2007 KNOWLEDGE FOR GENERATIONS

www.interscience.wiley.com

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

RECEIVED
JUL 24 2008
Levy Library

CONTENTS
VOL 75, NO 2

Introduction

63 Editorial: Learning from Painful Experiences
Penny A. Asbell

65 Theme Introduction: The 9/11 World Trade Center Disaster: Past and Future
J. Howard

Theme Articles

67 WTC Medical Monitoring and Treatment Program: Comprehensive Health Care Response in Aftermath of Disaster
Jacqueline M. Moline, Robin Herbert, Stephen Levin, Diane Stein, Benjamin J. Luft, Iris G. Udasin and Philip J. Landrigan

77 Epidemiologic Research on Man-made Disasters: Strategies and Implications of Cohort Definition for World Trade Center Worker and Volunteer Surveillance Program
David A. Savitz, Rachael T. Oxman, Kristina B. Metzger, Sylvan Wallenstein, Diane Stein, Jacqueline M. Moline and Robin Herbert

89 Upper and Lower Respiratory Diseases After Occupational and Environmental Disasters
David J. Prezant, Stephen Levin, Kerry J. Kelly and Thomas K. Aldrich

101 Potential for Diffuse Parenchymal Lung Disease after Exposures at World Trade Center Disaster Site
Jaime Szeinuk, María Padilla and Rafael E. de la Hoz

109 Standardization of Spirometry in Assessment of Responders Following Man-made Disasters: World Trade Center Worker and Volunteer Medical Screening Program
Paul Enright, Gwen Skloot and Robin Herbert

115 Mental Health of Workers and Volunteers Responding to Events of 9/11: Review of the Literature
Corey B. Bills, Nancy A. S. Levy, Vansh Sharma, Dennis S. Charney, Robin Herbert, Jacqueline Moline and Craig L. Katz

129 Impact of the September 11 World Trade Center Disaster on Children and Pregnant Women
Philip J. Landrigan, Joel Forman, Maida Galvez, Brooke Newman, Stephanie M. Engel and Claude Chemtob

135 Responder Safety and Health: Preparing for Future Disasters
Dori B. Reissman and John Howard

Special Clinical Features

143 Controversy (Co-Editors: Ira Nash, Anthony Squire)
Making Sense of ENHANCE: Ezetimibe (Zetia) Lowers LDL Cholesterol But Doesn't Decrease Carotid Intima-media Thickness
Donald Smith

Continued overleaf

Contents continued . . .

149 Evidence-based Medicine (Editor: Thomas McGinn)
Deciding Whether To Screen Patients for Abdominal Aortic Aneurysm
Thomas McGinn and Deborah Korenstein

153 Ethics (Editor: Rosamond Rhodes)
Capacity Assessment and Intervention in 56-yr-Old Man with Schizophrenia
Alan Schlechter

157 New Drugs and the FDA (Editor: Fred Lublin)
Neurotherapeutics in Multiple Sclerosis: Novel Agents and Emerging Treatment Strategies
Tracy DeAngelis and Fred Lublin

168 Prismatic Case (Editor: Maria New)
Report of Fertility in Woman with Predominantly 46,XY Karyotype in Family with Multiple Disorders of Sexual Development: Review of Prismatic Case
Miroslav Dumic, Karen Lin-Su, Natasha I. Leibel, Srecko Ciglar, Giovanna Vinci, Ruzica Lasan, Saroj Nimkarn, Jean D. Wilson, Ken McElreavey and Maria I. New

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

Editor-in-Chief

Penny A. Asbell

Deputy Editors

Dennis Charney
Maria New
C. Warren Olanow

Associate Editors

Stuart Aaronson
David H. Adams
Erwin Bottinger
Eric Hollander
Ravi Iyengar
Mary Klotman
Paul Klotman
Ihor R. Lemischka
Fred Lublin
Michael L. Marin
Thomas G. McGinn
Ira Nash
Rosamond Rhodes
Hugh Sampson
Mary Sano
Anthony Squire
Fred Suchy
Savio Woo

Editorial Board

William J. Bremner	University of Washington
Thomas M. Coffman	Duke University
Dwight Evans	University of Pennsylvania
Gabriel Haddad	University of California, San Diego
Heidi E. Hamm	Vanderbilt University
Matti Häyry	University of Manchester
Eve J. Higginbotham	Morehouse School of Medicine
Katherine A. High	University of Pennsylvania
Tony Hope	University of Oxford
John H. Krystal	Yale University
Raju Kucherlapati	Harvard Medical School
Robert J. Lefkowitz	Duke University
Robert Lisak	Wayne State University
George Lister	University of Texas
Philip A. Marsden	University of Toronto
Charles Nemeroff	Emory University
Scott Rauch	Harvard Medical School
Bernard Roizman	University of Chicago
Alan F. Schatzberg	Stanford University
Inder Verma	The Salk Institute
Jeffrey A. Whitsett	University of Cincinnati
Allan Wolkoff	Albert Einstein College of Medicine
Noam Zohar	Bar-Ilan University

Guest Co-editors

Robin Herbert, MD   Jacqueline Moline, MD

Dirs., WTC Medical Monitoring and Treatment Program at Mount Sinai Medical Center
(Dr. Herbert, Consortium Data and Coordination Center; Dr. Moline, Clinical Center)

Managing Editor

Stephen Cohen
stephen.cohen@mssm.edu
212-241-6304

Publisher

John Wiley & Sons
Jennifer English, Sr. Editor
jenglish@wiley.com

Answering the Call

The cover photo depicts one of the thousands of men and women who answered the call of duty after disaster struck the towers of the World Trade Center on September 11th. Among them were traditional first responders, construction, transportation and communications workers, and volunteers, all working in a landscape of utter devastation under conditions fraught with danger, including unstable footing, fires, smoke and toxic dust. To all of them this issue of the *Mount Sinai Journal of Medicine* is gratefully dedicated.

Acknowledgements by Guest Editors, Drs. Jacqueline Moline and Robin Herbert

In responding to the September 11, 2001 attacks on the World Trade Center, the medical community has faced an enormous challenge, requiring programs both to provide care for WTC responders and initiate research into the human health consequences of such attacks. In helping to direct these programs, we have been greatly privileged to work with extraordinary people from around the globe. For their support of the World Trade Center Medical Monitoring and Treatment Programs and their contributions to this special edition of the *Mount Sinai Journal of Medicine*, we would like to thank Drs. John Howard, Michael Galvin, and Sherry Baron of the National Institute of Occupational Safety and Health; Dr. James Melius, Chair of the WTC MMTP Steering Committee, and Drs. Philip Landrigan, Dennis Charney and Kenneth Davis, steering committee members; and legislators and the labor leaders who have been unswerving in their efforts to ensure ongoing services for WTC responders. We also wish to thank staff and patients of the World Trade Center Medical Monitoring and Treatment Program and the many labor, community and volunteer groups that have provided assistance. Special thanks to the New York Times Foundation, American Red Cross Liberty Fund, The September 11th Recovery Program, The Bear Stearns Charitable Foundation, The September 11th Fund, The Robin Hood Foundation Relief Fund, and many others. Finally, a particular thanks to the team who pulled this special edition together—Erica Hodgman, Elspeth Call and Rachael Oxman—and to the Mount Sinai School of Medicine for its commitment to this work over many years.

Themes for future issues of *MSJM*, each shown with Guest Editor(s)

- Diabetes type 1 (Derek LeRoith)
- Personalized medicine (Erwin Bottinger)
- Advanced lung disease (Maria Padilla)
- Hospital medicine (Andrew Dunn)
- Anxiety and mood disorders (Dennis Charney and Sanjay Mathew)
- Alzheimer's disease (Nilos Robakis)
- Climate change and human health (Michael McCally and Paul Landrigan)
- Liver disease (Paul Martin)
- Academic diversity (Gary Butts)
- Emerging pathogens (Mary Klotman)
- Medical education (David Muller)
- Drug discovery (Ravi Iyengar)

©2008 **Mount Sinai School of Medicine.** All rights reserved. No part of this publication may be reproduced in any form or by any means, except as permitted under section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the publisher, or authorization through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, Phone: (978) 750-8400; Fax: (978) 750-4770. Such requests and other permission inquiries should be addressed to the Permissions Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; Phone: (201) 748-6011; Fax: (201) 748-6008; E-mail: permreq@wiley.com.

Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine (Print ISSN: 0027-2507; Online ISSN: 1931-7581 at Wiley InterScience, www.interscience.wiley.com) is published twelve times per year for the Mount Sinai School of Medicine by Wiley Subscription Services, a Wiley Company, 111 River Street, Hoboken, NJ 07030.

Send subscription inquiries to: Subscription Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-6645.

Offprint sales and inquiries should be addressed to the Reprint Billing Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-8776. All other inquiries should be addressed to the Customer Service Department at (201) 748-6645.

Individual subscription price: $95.00 in North America, $120.00 outside North America. Payment must be made in US dollars drawn on a US bank. Claims for undelivered copies will be accepted only after the following issue has been delivered. For more subscription prices and information, please visit www.mountsinaijournal.org.

Change of address: Please forward six weeks prior to move: present mailing label with change of address. Please enclose a copy of the mailing label or cite your subscriber reference number in order to expedite handling. Missing copies will be supplied when losses have been sustained in transit and where reserve stock permits. Send claims in care of: *Mount Sinai Journal of Medicine*, c/o John Wiley & Sons, Inc., Attn: Journals Admin Dept. UK, 111 River Street, Hoboken, NJ 07030.

Cancellations: Subscription cancellations will not be accepted after the first issue has been mailed. Periodicals postage paid at Hoboken, NJ and at additional mailing offices.

Postmaster: Send address changes to *Mount Sinai Journal of Medicine*, Subscription Distribution, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Printed in the United States of America.

Advertising inquiries should be addressed to: Advertising Sales Manager, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Phone: (201) 748-6921; Fax: (201) 748-6207. Advertising in this journal is published with a caveat emptor understanding. The Publisher, Editors, and members of the Editorial Board do not imply warranty of products, nor quality, validity, or approval of education material offered by such advertising.

Indexed by: Biological Abstracts® (Thomson ISI), Chemical Abstracts Service/SciFinder (ACS), Current Contents®/Clinical Medicine (Thomson ISI), EMBASE/Excerpta Medica (Elsevier), Index Medicus/MEDLINE/PubMed (NLM), Science Citation Index Expanded™ (Thomson ISI), SCOPUS (Elsevier), Web of Science® (Thomson ISI).

This journal is printed on acid-free paper.

Disclaimer: All articles published, including but not limited to original research, reviews, clinical notes, clinical pearls, prismatic cases, clinician's corner, editorials, reports, and letters, represent the opinions and views of the authors and do not reflect any official policy or medical opinion of the Mount Sinai School of Medicine, the institutions with which the authors are affiliated, or the Publisher unless this is clearly specified. Articles published herein should not be relied upon as recommending or promoting a specific method, diagnosis, or treatment by physicians for any particular patient or disorder. Although the Editors and Publisher believe that drug selections and dosages and the specifications and usage of equipment and devices as set forth herein are in accord with current recommendations and practice at the time of publication, they accept no legal responsibility for any errors or omissions, and make no warranty, express or implied, with respect to material contained herein.

Mount Sinai Journal of Medicine 75:63–64, 2008

EDITORIAL

Learning from Painful Experiences

Penny A. Asbell

Mount Sinai Medical Center, New York, NY

This issue of the *Mount Sinai Journal of Medicine* is devoted to an in-depth presentation of the healthcare lessons learned from the World Trade Center (WTC) disaster. We all know the difficulties of determining the statistical significance of findings when the denominator–here the size of the affected population exposed to health dangers related to the disaster–is not known, but the effort must be made if we are to maximize our knowledge of the health consequences of the WTC disaster. The lessons learned are important not only for addressing the needs of individuals affected by this disaster but also for planning our responses to other disasters in the future.

Mount Sinai Medical Center took a leading role in initiating medical screening procedures to document health problems related to the WTC disaster and also developed treatment and ongoing surveillance programs for responders. Our guest editors, Dr. Robin Herbert and Dr. Jacqueline Moline, continue to maximize the evidence-based information from these efforts while at the same time providing care for volunteers and responders to the WTC disaster who have continuing health problems.

This issue of the *Mount Sinai Journal of Medicine* reviews some of the information gathered to date on health consequences of the WTC disaster and the plans to improve disaster response in the future to protect the health of responders:

- Dr. John Howard, Director of the National Institute for Occupational Safety and Health, in his introduction to this issue calls our attention to the unprecedented scale of the disaster and the planned and spontaneous response that it evoked and stresses the need for further scientific study to document the physical and mental health impact arising from exposure to the disaster.
- Dr. Robin Herbert and Dr. Jacqueline Moline, both at Mount Sinai Medical Center, and their coauthors describe the comprehensive healthcare response that they developed at Mount Sinai Medical Center to address the heterogeneous group of traditional and nontraditional responders involved in the rescue, recovery, cleanup, and restoration in the aftermath of September 11th. They point out that although most disasters involve a single known exposure among a well-described cohort, the WTC disaster presented us with a situation in which there were limited data on both exposure and health consequences and no easy access to information that would identify exposed individuals. For the future, they call for the development and improved utilization of community organizations to implement better and more comprehensive disaster-response programs.
- Dr. Savitz and his coauthors discuss the difficulties in reaching valid inferences regarding elevated health risks when enrollment in the surveillance program is selective.
- Dr. Dori Riesmann, from the Office of the Director of the National Institute for Occupational Safety and Health, reviews how regional and national organizations can plan for the unplanned disasters that occur. Such planning will lead to improved safety and health and to better preparedness.
- Dr. Rafael E. de la Hoz and his coworkers review the current evidence linking exposure to the WTC disaster with lung disease and conclude that it is possible that diffuse parenchymal lung disease may be associated with severe exposure to dust and smoke at the WTC disaster site.
- Dr. Robin Herbert and her coauthors describe their efforts to standardize testing for lung disease using

Address Correspondence to:

Penny A. Asbell
Editor in Chief
Mount Sinai Journal of Medicine
Mount Sinai Medical Center
New York, NY
Email: penny.asbell@mssm.edu

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20036

© 2008 Mount Sinai School of Medicine

spirometry, and they demonstrate that this simple test can provide a standard evaluation of lung disease even when used in multiple examination sites. Such standardized tests can greatly help in providing evaluable data on lung disease following hazardous health exposure. They recommend this approach for postdisaster pulmonary evaluations.
- Dr. David Prezant, Chief Medical Officer of the Fire Department of the City of New York, and his coauthors illustrate in detail the consequences of chronic upper and lower respiratory inflammation as a result of exposure to the WTC disaster and describe in detail specific respiratory syndromes that have been observed.
- Dr. Craig Katz and his coauthors review and analyze the published material on mental health issues affecting disaster workers at the WTC site and note other significant mental health problems in addition to posttraumatic stress disorder.
- Dr. Phil Landragan describes findings in women and children exposed to the aftermath of the WTC disaster and expresses concern for potential delayed health consequences in this vulnerable population.

This issue of the *Mount Sinai Journal of Medicine* also introduces our new feature on utilization of evidence-based medicine in the practical care of patients. Dr. Tom McGinn is a world-renowned teacher who is devoted to ensuring that scientific information is correctly interpreted so that it can be wisely used for practical patient care. Regardless of one's specialty, we know that this feature will help all of us in clinical practice to better interpret the deluge of information that we receive on a daily basis.

IN MEMORIAM

We regret to note the passing of an enthusiastic supporter of the relaunch of the *Mount Sinai Journal of Medicine*, Dr. Judah Folkman, a member of our board of editors. He epitomized the bench-to-bedside and back again opportunities for real breakthroughs in patient care. He was committed to a lifetime of persistence and innovation in scientific research based on evidence-based information. We met nearly 2 decades ago when I suggested that we apply his work to eye disease, specifically corneal neovascularization, and although our work was not successful, I think that it sparked his interest in applying his efforts to eye diseases such as macular degeneration. He was a good man, and he will be sorely missed.

DOI:10.1002/MSJ

THEME INTRODUCTION

The 9/11 World Trade Center Disaster: Past and Future

John Howard, MD

National Institute for Occupational Safety and Health,
Centers for Disease Control and Prevention, Washington, DC
World Trade Center Health Programs, US Department of Health and Human Services, Washington, DC

On September 11, 2001, terrorists hijacked four passenger airliners and flew one into the North Tower and another into the South Tower of New York City's World Trade Center. Shortly thereafter, both towers collapsed. Major trauma, burns, overwhelming smoke inhalation, and falls from great heights killed more than 2000 tower occupants, 147 airline passengers (in addition to the hijackers), 343 firefighters from the New York City Fire Department, 23 officers from the New York City Police Department, and 37 officers from the Port Authority of New York and New Jersey Police Department.

The disaster triggered an unprecedented emergency response involving hundreds of early arriving rescue workers and volunteers–both previously trained volunteers and "spontaneous" volunteers. Thousands of later arriving rescue responders and volunteers engaged in search and rescue activities to free those who were trapped in the rubble of the collapsed towers. In the months immediately following September 11th, tens of thousands of additional responders and volunteers from all over the United States worked tirelessly to retrieve human remains; to recover and clean up the site itself and the streets, residences, and commercial buildings in Lower Manhattan; to transport an enormous amount of debris to barge stations located along the west side of Lower Manhattan; to receive, handle, and, in some cases, shift through the debris at the Staten Island Landfill; and to perform forensic examinations of human remains at the New York City Medical Examiner's Office.

The collapse of the towers immediately generated a massive cloud of dust. Persistent smoke resulted from fires that burned until late December 2001. The dust generated by the collapse and the smoke from the fires contained numerous chemical and physical agents known to be hazardous to human health. Responders and volunteers who were actively engaged in the nearly year-long rescue, recovery, and clean-up activities and several hundred thousand people who live, work, and attend school near the World Trade Center site were exposed to the dust, smoke, and concomitant hazardous agents from the place that became known as *Ground Zero*.

Not unexpectedly, some early responders developed acute respiratory conditions from massive dust inhalation, and some experienced significant stress reactions. In the months that followed the disaster, many responders and some Lower Manhattan residents, students, and building occupants noted new or worsened aerodigestive and mental health problems. In response to these early findings, medical screening of firefighters, emergency medical services personnel, and other responders and volunteers was begun along with recruitment of all exposed persons to the World Trade Center Health Registry. More recently, ongoing medical monitoring and treatment programs for responder and nonresponder populations have been established by federal and city governments to care for those whose physical and mental health has been adversely affected by their exposure to the World Trade Center disaster.

Address Correspondence to:

Jacqueline Moline
Department of Community and
Preventive Medicine
Mount Sinai School of Medicine
New York, NY
Email: jacqueline.moline@
mssm.edu

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20037

© 2008 Mount Sinai School of Medicine

Despite the progress that has been achieved to date in addressing the needs of World Trade Center–affected populations, more still needs to be done. One area that needs more attention is the scientific study of the physical and mental health impact of exposure to the World Trade Center disaster. I thank the editors of the *Mount Sinai Journal of Medicine* for dedicating a special edition of the journal to scientific papers that help us better understand that impact.

As the new World Trade Center towers rise above the skyline of New York City in the years ahead, let us not forget those who gave their lives and risked their health in their heroic response to the events of September 11, 2001.

DOI:10.1002/MSJ

WTC Medical Monitoring and Treatment Program: Comprehensive Health Care Response in Aftermath of Disaster

Jacqueline M. Moline, MD, MSc,[1] Robin Herbert, MD,[1] Stephen Levin, MD,[1] Diane Stein, BS,[1] Benjamin J. Luft, MD,[2] Iris G. Udasin, MD[3] and Philip J. Landrigan, MD, MSc[1]

[1] Mount Sinai School of Medicine, New York, USA
[2] SUNY at Stony Brook, New York, USA
[3] University of Medicine & Dentistry of New Jersey, USA

ABSTRACT

The attack on the World Trade Center (WTC) on September 11th, 2001 exposed thousands of individuals to an unprecedented mix of chemicals, combustion products and micronized building materials.

Clinicians at the Mount Sinai Irving Selikoff Center for Occupational and Environmental Medicine, in partnership with affected stakeholder organizations, developed a medical screening program to evaluate the health status of workers and volunteers who spent time at the WTC site and thus sustained exposure in the aftermath of September 11th. Standardized questionnaires were adapted for use in this unique population and all clinicians underwent training to ensure comparability.

The WTC Worker and Volunteer Medical Screening Program (MSP) received federal funding in April 2002 and examinations began in July 2002. The MSP and the follow up medical monitoring program has successfully recruited nearly 22,000 responders, and serves as a model for the rapid development of a medical screening program to assess the health of populations exposed to environmental hazards as a result of natural and man-made disasters.

The MSP constitutes a successful screening program for WTC responders. We discuss the challenges that confronted the program; the absence of a prior model for the rapid development of a program to evaluate results from mixed chemical exposures; little documentation of the size of the exposed population or of who might have been exposed; and uncertainty about both the nature and potential severity of immediate and long-term health effects. ***Mt Sinai J Med* 75:67–75, 2008.** © *2008 Mount Sinai School of Medicine*

Key Words: World Trade Center, screening, exposure, responders, workers, occupational.

Address Correspondence to:

Jacqueline M. Moline
Mount Sinai School of Medicine
Department of Community &
Preventive Medicine
New York, NY, USA
Email: jacqueline.moline@mssm.edu

The attack on the World Trade Center (WTC) on September 11, 2001 exposed thousands of individuals to an unprecedented mix of aerosolized chemicals, combustion products and micronized building materials. For many, exposures began on September 11th, 2001 and continued in the weeks and months thereafter while they toiled in rescue and recovery work; others arrived later but also risked exposure due to fires, which burned into December 2001, and resuspension of settled dust. There was an overwhelming response from thousands of workers and volunteers to provide needed rescue and recovery operations. Those who responded included traditional first responders, such as firefighters, law enforcement officers and emergency service workers, and others who conducted a range of job tasks. See Table 1 for sample job titles. Owing to the overwhelming scale of the disaster, conditions at the site were hazardous and essentially unregulated, particularly in the first

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20022

© 2008 Mount Sinai School of Medicine

Table 1. *Job Functions of World Trade Center Responders.*

Traditional
Emergency service workers
Federal disaster responders
Firefighters
Law enforcement
Urban search and rescue (USAR)
Nontraditional Workers
Building cleaners
Building trades
Civil service workers
Communication workers
Counselors
Engineers
Environmental assessment workers
Media
Mortuary workers
Nonemergency health care workers
Pastoral care
Public officials
Sanitation workers
Transport workers
Utility workers
Veterinarians
Volunteers

days after September 11th. Few systematic rosters were maintained of those working in and around the site; respiratory protection was not always available[1] and, even when available, was not suitable for the prevailing conditions.[1]

Given the magnitude of the tower collapses and continuing fires, it was evident that many people were at risk for significant exposures to WTC dust, products of combustion, and debris. The potential for exposure was significant for all those individuals present at the time of the building collapse, especially for those caught "in the dust cloud", as well as for other workers who arrived at the site later on September 11th or in the subsequent days and weeks. Additionally, responders were exposed to a myriad of events and sights known to induce psychological trauma.

Although there are models for acute and mental health care of traditional responders involved in disaster response,[2,3] there is far less information in the literature on postdisaster medical monitoring and treatment for traditional and nontraditional responders with exposures to an array of toxic materials such as the complex mix of contaminants released after the 9/11 attacks. This paper describes the establishment of a comprehensive program to identify and provide medical monitoring and treatment to a heterogeneous group of traditional and nontraditional responders involved in rescue, recovery, clean up and restoration of services in the aftermath of the attacks on the WTC.

INITIAL MEDICAL CONCERN AND RESPONSE

The attack on and the collapse of the WTC towers and the fires at the site that continued for several months generated considerable concern among a variety of medical specialists that health consequences could result in individuals exposed at the site. In the immediate aftermath of September 11th, concerns focused on acute health conditions, such as musculoskeletal injuries, and acute respiratory conditions, such as reactive airways disease, associated with dust and fume exposure and the psychological implications from the disaster exposures. In addition to early, acute health concerns,[4] there were concerns regarding the particular vulnerability of the great number of "nontraditional" disaster responders without prior training in disaster response. Many of these responders constituted a vulnerable population from a health and safety perspective. There was also heightened concern of the risk for chronic and late emerging conditions related to WTC exposures.

The Mount Sinai Irving J. Selikoff Center for Occupational and Environmental Medicine (COEM) began evaluating individuals with WTC exposures within weeks of the tragedy. The Selikoff Center is a university-based occupational health center of excellence that had been supported by the New York State Department of Health for nearly 20 years prior to 9/11.[4] It has vast experience in examining workers exposed to asbestos, lead, silica, and other industrial pollutants. On the basis of anecdotal clinical reports from various sources, and because of the limited information about the nature of exposures and their health effects, officials from the National Institute for Occupational Safety and Health (NIOSH) initiated regular conference calls to facilitate communication among health care providers on "the front lines" of caring for responders. This initiated a best practice model in which initial clinical observations from providers treating similarly exposed populations were shared, with the goal of identifying possible disease trends and sharing diagnostic treatment modalities.

During the same time period, similar concerns began to emerge from discussions with local union representatives, who were disturbed by the health conditions among their constituents. During the fall of 2001, the unions, partnering with COEM physicians and others, raised awareness among locally and nationally elected officials and agency representatives of the need for health monitoring of WTC exposed individuals. Union representatives and others having direct contact with responders recognized that there was a very broad spectrum of tasks and jobs required to mount an effective rescue and recovery operation. Not only were individuals working directly on "the pile" exposed, but so were thousands of others who worked to restore essential services around the WTC site. In addition, building cleaners and those carting and sorting through debris, were at potential risk of developing health problems.

There was an evident need for a medical monitoring program that could meet the needs of the diverse WTC responder population. In early 2002, federal funding was committed to establish and conduct a medical monitoring program for WTC rescue and recovery workers not from the fire department of New York (FDNY). FDNY responders [firefighters, fire officers, and the FDNY/Emergency Medical Service (EMS) workers] were excluded from the program since they had access to an equivalent program that had been rapidly implemented–enabled by a robust preexisting clinical and data management infrastructure–by the FDNY in September 2001 through the department's Bureau of Health Services. This was expanded to include volunteer FDNY and FDNY EMS retirees through federal funds in December 2001. In April 2002, the Center of Disease Control and Prevention (CDC), through NIOSH, awarded funding to the Mount Sinai COEM for program planning to establish the WTC Worker and Volunteer Medical Screening Program (WVMSP). The contractors were instructed, however, that the source of funding (Federal Emergency Management Agency, FEMA) precluded the provision of treatment as well as data collection for scientific purposes: funding was limited strictly to the provision of medical monitoring examinations. An intensive 3-month program design, planning, and implementation period commenced in April 2002.

CONSTRUCTING THE MOUNT SINAI COEM WTC SCREENING PROGRAM

There was little precedent for setting up a health monitoring program of this scale for such a heterogeneous group of exposed individuals. Surveillance programs for occupational exposures have traditionally been designed to follow health outcomes related to a single known exposure among a well described cohort.[5–9] The creation of the screening program for WTC responders posed several challenges:[1] There was limited data on both exposures and health consequences in the initial aftermath of the disaster;[2] there was no single entity to coordinate identification, recruitment and scheduling of responders from a wide range of worksites; and[3] there was no systematic roster that had been kept of those who had responded. Thus, not only did program design and implementation proceed simultaneously, but the establishment of the program required the creation of a new model of postdisaster surveillance–one that provided examinations to meet the clinical needs of a broad mix of responders, who were linked by exposure and by concern about potential and unknowable future health effects.

The Occupational Safety and Health Administration (OSHA) defines medical screening as "a method for detecting disease or body dysfunction before an individual would normally seek medical care. Screening tests are usually administered to individuals without current symptoms, but who may be at high risk for certain adverse health outcomes"[10] due to occupational exposures. OSHA also dictates standard requirements (including types of examinations, frequency of examination and exposure tests) for screening of workers with exposure to certain single toxic agents.[11] There are no specific guidelines for addressing multiple exposures outside of hazardous materials, nor are there guidelines for civilian responders in postdisaster monitoring. There are, however, some publications in the medical literature addressing the guiding principles and objectives of occupational screening. Halperin *et al.* state that "the goals of screening are assumed to be 1) the early detection of disease and therapy of disease; 2) the evaluation of adequacy of exposure control and other means of primary prevention; 3) the detection of previously unrecognized health effects suspected on the basis of toxicologic and other studies; and 4) the suitable job placement".[12] Notably, these goals were structured with the standard occupational work site in mind.

In the unique setting of the WTC disaster response, there were numerous factors that contributed to the "atypical" work environment. There was a tremendous urgency of working conditions. There was a loss of communication due to damage to telephone systems, and no ability to communicate at the site except by direct vocalization (i.e. two-way radios did not function). There was the uncertainty about the status and safety of family members, friends, and colleagues. There were concerns about ongoing structural instability, fires and explosions, and concern that another attack was imminent, and that there were potential exposures to unknown substances. There was the presence of multiple authorities from local, state, and the federal government who gave conflicting information. Access to the WTC site in the early days was uncontrolled and there was lack of standardized training requirements. Workers and volunteers responded, working for extended work shifts and sleeping onsite. The availability of appropriate safety supplies was initially limited. These factors made application of preexisting occupational medical surveillance programs difficult, if not inappropriate, to a disaster worker population.

Programmatic structure and implementation are less defined in the literature than the overt goals of screening. Baker and Matte delineated steps for design and implementation of a program, as summarized in Table 2.[13,14] Their design highlights the importance of developing an examination protocol that is focused and standardized with respect to exposure assessment, examination administration, and data collection. These principles were applied to the design of the MSP.

The MSP was envisioned as a multifaceted medical program to address outcomes and exposures that were ill defined for an overwhelmingly large heterogeneous population. There were clear goals for the MSP, which included:

- Identification of individuals who had sustained exposures of the WTC site during rescue and recovery operations (including restoration and provision of vital services and removal of debris)
- Provision of clinical assessments for exposed individuals to identify persistent WTC-related medical conditions, using a consortium of clinical sites in the NY/NJ area that had known expertise in occupational health surveillance
- Provision of coordinated referral for follow-up clinical care for affected individuals
- Education of individuals about their exposures and the associated health risks, as well as advice on available benefit and entitlement programs
- Establishment of a "baseline" clinical status for individuals exposed at or near "Ground Zero" for the purpose of comparison with future clinical assessments, particularly for diseases with longer latency (In contrast to the FDNY Bureau of Health Services, there was no pre-WTC exposure health monitoring information for the majority of anticipated participants in the MSP.)

Table 2. *Steps in Designing and Implementing an Occupational Health Surveillance Program.*

Assessment of workplace hazards
Identification of target organ toxicities for each hazard
Selection of test for each "screenable" health effect
Development of action criteria
Standardization of data collection process
Performance of testing
Interpretation of test results
Test confirmation
Determination of work status
Notification
Diagnostic evaluation
Evaluation and control of exposure
Record keeping
(From Baker and Matte)

Rapid Program Development

Given the short time period within which the program had to be developed, multiple critical components of the program had to be developed simultaneously. They included the following:

- identification of a group of experts to assist in development of examination content and protocols;
- development of an examination protocol and standardized instruments;
- establishment of exposure-based eligibility criteria.

An Executive Steering Committee (ESC) was convened to oversee aspects of program planning and policy development. Representatives from organized labor, employer groups, all clinical consortium centers, and experts from multiple fields, including occupational medicine, pulmonary medicine, psychiatry, epidemiology, industrial hygiene, and health education served on the ESC and guided the development of the program. The ESC provided a forum for responder representatives to participate in all aspects of the program, from initial cohort identification and development of eligibility criteria. It provided program guidance on aspects of the protocol and feedback on its effectiveness in reaching stakeholders. Through monthly meetings with program leadership from each participating center, the ESC was able to be informed of issues confronting responders.

To facilitate rapid development of the program, four core areas were identified as essential elements of the "screening program": (1) outreach and exposure assessment; (2) administration; (3) medical evaluation; and (4) data management. Three working groups were convened to develop protocols for the program: medical; exposure assessment; and outreach/education.

Outreach and Exposure Assessment

The outreach/education and exposure working groups communicated frequently to determine methods to identify the types of workers involved in rescue, recovery, restoration of services, and cleanup, as well as means to identify the most heavily exposed potential participants. While many unions and employers had some records of who participated in rescue and recovery operations, there was no comprehensive database of those involved, nor, in most cases, accurate employment records. Additionally, there was overlap between those exposed in the line of duty and those exposed as a result of volunteer activities: in the postdisaster days, responders frequently worked for extremely long hours, with many employees continuing to work in a volunteer capacity after their official shift was over. Many also slept in the vicinity of the disaster. In addition, in the early days, as countless individuals simply arrived to volunteer their services, their presence was not always "officially" recorded.

One of the critical questions facing the screening program working groups was the uncertainty about the demand for clinical evaluations among exposed responders. The actual number of responders who would be eligible for the non-FDNY program was unclear when the program began, estimates ranged from 15,000 to 30,000–40,000 workers and volunteers. Because there were inherent limits on the number of participants who could ultimately be accommodated in the program, exposure based eligibility criteria were adopted to ensure that individuals with the greatest potential need, i.e. greatest exposures, could participate in the program. Over time, as the clinical activities of the screening program began and outreach activities were underway, these criteria were revised to accommodate individuals who had arrived later in the month of September or had fewer total hours of WTC exposures. All changes to eligibility criteria were approved by the ESC. The final eligibility criteria for participation in the screening program defined two categories of eligible individuals as follows:[1] the responder must have operated as a rescue, recovery, debris-cleanup or related support services worker, and/or volunteer who was located in lower Manhattan (south of Canal Street) and/or the Staten Island Landfill and/or the barge loading piers for a minimum duration of 4 hr on September 11th–14th, 2001 OR at least 24 hr in the month of September OR at least 80 hr for the months of September, October, November, and December combined;[2] the responder was an employee of the Office of the Chief Medical Examiner (OCME) or other morgue worker who was involved in the examination and processing of human remains OR the responder was a worker in the Port Authority Trans-Hudson Corporation tunnel for a minimum duration of 24 hr between September 11th, 2001 and July 1st, 2002 OR the responder was a vehicle-maintenance worker within the requisite time frames who was exposed to WTC debris while retrieving, driving, cleaning, repairing, and maintaining contaminated vehicles.

Eligibility interviews were conducted by a trilingual phone bank that fielded 138,982 calls during the course of the screening program. The phone bank also became a clearinghouse for referrals to programs that offered some assistance to callers who were deemed ineligible for the MSP.

An outreach unit was established and staffed by personnel experienced in occupational health and familiar with key organizations, primarily labor unions. An intensive outreach campaign was conducted that included development of written recruitment/outreach materials, participation in hundreds of meetings with unions and other responder organizations, public service announcements in English and Spanish, and an extensive media campaign and other strategies to increase awareness of the program throughout responder communities. Outreach staff worked directly with leadership from nearly 70 unions, which were pivotal in identifying responders, to customize outreach activities based on the unions' experiences in communicating with their members. In some instances speakers were provided at union meetings. Other strategies included providing articles for union newsletters, sending letters from the union president directly to members, assisting with telephone messages to members from union leadership, recording messages to be played on hold buttons at union headquarters, and going to work locations with union leadership to inform workers about the program.

Educational materials were developed for workers to inform them of potential hazards both at the WTC site and in their current and previous occupations, and were distributed during the exposure assessment interview. A detailed interviewer-administered exposure assessment instrument was developed to obtain pre- and post-September 11th related duties. Information about locations at or around the WTC site, Staten Island landfill or barges carrying debris was collected, along with dates of services and hours worked. Because specific exposure data is absent, time spent and location at the WTC site served as surrogates for overall exposure.

Administration

On the basis of specific advice from NIOSH, Mount Sinai developed a clinical consortium of regional partners, with long standing occupational health expertise, which would conduct examinations at geographic locations most convenient to responders. Consortium partners included: the Bellevue/NYU Occupational and Environmental Medicine Clinic, State University of New York at Stony Brook/Long Island Occupational and Environmental Health Center, the Center for the Biology of Natural Systems at Queens College in N.Y., and the Clinical Center of the Environmental and Occupational Health Services Institute at UMDNJ–Robert Wood Johnson Medical School in New Jersey. For those responders and volunteers who came from outside the NY/NJ region, clinical services were coordinated nationally by the Association of Occupational and Environmental Clinics. The lead physician at each of the clinical consortium sites was a member of the ESC and served on the medical working group.

Medical Evaluation

The medical working group (MWG) was established to develop the clinical protocol for a standardized occupational health screening examination intended to assess both physical and mental health. During the protocol development, it was recognized that information should be collected for a research database to underpin scientific assessment of the full impact of the disaster. The primary goal of the MSP however, was provision of clinical services and this was given priority when clinical protocols conflicted with the collection of research data.

Many elements of the clinical screening evaluation were modeled on asbestos screening examinations that had been performed at the Mount Sinai Irving J. Selikoff COEM for more than 30 years.[15,16] The clinical screening examination consisted of physical and mental health questionnaires, a standardized physical examination, pre- and postbronchodilator spirometry, complete blood count, blood chemistries, urinalysis and chest radiograph (see Table 3). Serum banking was performed at some clinical sites. Examination results were discussed with each patient in a face-to-face consultation with a physician at the end of the overall evaluation. All participants received mailed, written summaries of their evaluation. Referrals for further evaluation and treatment of physical and mental health conditions were also provided to participants, if indicated.

Self-administered and interviewer-administered physical and mental health questionnaires were developed using adaptations of standardized instruments whenever possible. For the physical health questionnaires, standardized questions were incorporated when feasible from the Bronchial Symptoms Questionnaire, European Community Respiratory Health Survey, NHANES III surveys, NIOSH Health Hazard Evaluations, and Sino-Nasal Outcome Test.[17–19,22,23] The self-administered mental health questionnaire was a combination of several standardized questionnaires, including the General Health Questionaire-28, Post-Traumatic Stress Disorder Symptom Checklist, Patient Health Questionnaire, CAGE Questionnaire, and Sheehan Disability Scale.[20,21,24–26] Participants who met threshold criteria for possible mental health impairment based on their questionnaire responses were referred for a clinical mental health evaluation. Additionally, any patients who acknowledged suicidal ideation or substantial disability, or those who requested to speak to a mental health professional, were referred for clinical evaluation that day. A critical feature of the screening program was the integration of the mental health component with the physical health program. The self-administered physical and mental health questionnaires were distributed together, in an effort to decrease stigma attached to the completion of mental health questions. Furthermore, the clinical mental health evaluation was considered an equal element of the overall clinical component, with mental health services provided at the same time as the physical health evaluation.

All clinicians participating in the screening program underwent standardized training at in-person training sessions with experts in interviewing techniques and specialists in the evaluation of the nasopharynx and in interpretation of spirometry results. For clinicians who were evaluating responders outside the NY/NJ area, a training video was developed, consisting of taped lectures and demonstrations by these experts. In addition, all pulmonary function tests given via the NY/NJ consortium were performed using the same type of spirometer and all technicians were trained by the same experts in

Table 3. *Clinical Components of Screening Examination.*

Self-Administered Questionnaires
  I. Self-Administered Mental Health questionnaire (17–21)
  II. Self-Administered Physical Health questionnaire (14–16,22,23)
Interviewer-Administered Questionnaires
  III. Interview Administered Medical questionnaire, IAMQ (14–16,22,23)
  IV. Exposure Assessment Questionnaire
Examination
  V. Physical Examination
Diagnostic Tests
  VI. Pulmonary Function Test–EasyOne® Spirometer (24,25)
  VII. Labs (Blood Count, Blood Chemistries and urinalysis)
  VIII. Chest Radiograph
Examination Results
  IX. Physician Exit Interview (Immediate Letter–"Urgent Letter" for notification of abnormal laboratory or chest radiograph)
  X. Written Summary of Results (Final letter)

administering the test.[27] Emphasis was also placed on obtaining NIOSH certificates for pulmonary technicians. These training sessions were a critical aspect of staff development and training and ensured uniformity of testing within the program as a whole.[28] It also provided training for referral for further care of conditions felt to be WTC-related.

Almost 12,000 responders participated in the medical screening program between July 2002 and July 2004, including 650 individuals seen outside the NY/NJ region. Nearly 9,000 individuals were seen at Mount Sinai alone. Additional funding was obtained in July 2004 to continue to monitor the physical and mental health of the WTC responders. Since then over 12,000 additional responders have received initial monitoring examinations, increasing the total cohort size to approximately 22,000 responders.

Data Management Core Functions

The Data Management core was responsible for developing databases and entering data gathered within the screening program. However, insufficient funds were provided for data cleaning and data analysis.

Establishment of the Health Effects Treatment Program

As noted previously, the federal funding for the MSP was allotted for screening procedures and mental health services only, with no federal funding designated for clinical diagnostic services or treatment. The development and implementation of the WTC treatment program is beyond the scope of this discussion, however it should be noted that medical monitoring initiates medical responsibility to ensure that workers receive appropriate medical attention. From the outset, responders seen in the WTC WVMSP were able to receive needed mental health treatment at some clinical sites through separate federal and philanthropic funding.[29] Additionally, concomitant with the establishment of the MSP, program leaders began to seek funds to establish a treatment program for WTC responders. Philanthropic funding was solicited in early 2002 for appropriate treatment and support of responders suffering from physical, personal and socioeconomic repercussions of WTC exposure. Funding was received from a number of benefactors, enabling the inauguration of the WTC Health Effects Treatment Program (HETP) in January 2003.[30] A program for treatment of mental health conditions arising from or aggravated by the WTC disaster also received philanthropic funding in 2002 and the program began in 2003.[29] Federal funding for physical and mental health treatment services was provided in November 2006 to the WTC Medical Monitoring Program, which has subsequently been renamed the WTC Medical Monitoring and Treatment Program (MMTP).

DISCUSSION

The events of September 11th, 2001 were unprecedented not only in terms of national security, but also from a public health perspective. Despite its extensive network of hospitals and thousands of physicians, New York City was not immediately equipped to address the potential health risks to the thousands of responders who arrived at the WTC site in the days, weeks and months following the collapse of the towers. Given the unique nature of the events and subsequent hazards, there were no models upon which to structure the WTC Medical Screening/Monitoring Program (MS/MP).

The factors that make the WTC MS/MP unique among health monitoring programs include the following:

- provision of free, confidential, comprehensive, and highly standardized medical and mental health screening examinations for a highly heterogeneous group of citizens who responded to a disaster with exposures of unknown health consequences;
- Rapid development of a concomitant physical and mental health treatment program;

Utilization of a bio-psychosocial model of care, addressing the physical, mental health, and socioeconomic well being of the responders, with screening for psychological, social, and physical problems incorporated into the same clinical evaluation.

In order to accomplish the formidable task of providing comprehensive screening, monitoring and eventually treatment for the WTC responders, several critical features were essential to the program's success:

- Reliance on the preexisting infrastructure of the New York State Occupational Health Clinic Network–a unique system of Centers of Excellence in Occupational Health that had been in existence for nearly 20 years.
- Involvement and inclusion of responder representatives and other program constituents in the development, evaluation and governance of the program from its inception.
- Use of a comprehensive bio-psychosocial model to ensure that the breadth of responders' needs was addressed.

- Implementation of a constant feedback loop from clinicians, modeled on systems that track sentinel health events.

CONCLUSION

The horrific events of September 11th, 2001 provided the impetus for policy makers to review and strengthen procedures involving national security and other related issues. Public Health advocates must also use this event as a catalyst to examine disaster response. A large number of lessons have been learned as a result of this program that addressed health consequences suffered by thousands of WTC responders.

The overarching lesson is that postdisaster health screening, monitoring and treatment must be adapted to fit the needs of the population and hazards likely to be involved in future disasters.

Future disaster response programs should at a minimum provide the following:

- identification of physical and mental health disease in individuals;
- identification of socioeconomic needs of individuals;
- identification of disease patterns and socioeconomic need patterns among the constituent group;
- identification of resources for affected individuals to ensure the delivery of appropriate physical and mental health care;
- identification of resources for affected individuals to ensure that their basic economic needs are met

These programs will be most effective if they:

- Achieve early identification of the affected group
- Utilize a "community organizer" approach by involving the constituent population in the development, implementation and evaluation of the program
- Accommodate the heterogeneous nature of the American workforce. These accommodations must include evaluation of proper communication tools including variables such as language comprehension and literacy. Additional factors include knowledge of occupational health protections and general access to medical care.

In every disaster, hundreds and even thousands of civic-minded people immediately rush to the site to help. It is incumbent on us to ensure the health and well-being of these men and women by providing protection during their work and to provide medical monitoring where necessary, treatment upon conclusion of the work. The WTC MSP and MMTP provides a model for a program that has met these challenges.

We would like to acknowledge the contributions and support of the following individuals and organizations: J Howard, S Baron, M Galvin, R Fleming, V Sharma, C Katz, R Smith, S Markowitz, D Harrison, R Oxman, E Call, American Red Cross Liberty Fund, The September 11th Recovery Program, The Bear Stearns Charitable Foundation, The September 11th Fund, The Robin Hood Foundation Relief Fund, and many others.

ACKNOWLEDGMENTS

This work was supported by the Centers for Disease Control and Prevention and the National Institute for Occupational Safety and Health, grant C200-2002-00384 (Screening Program), grant 5U10 OH008232-04 (DCC), and grant 3U10 OH008225 (CC).

REFERENCES

1. Jackson BA, Peterson DJ, Bartis JT, et al. *Protecting Emergency Responders: Lessons Learned from Terrorist Attacks*. Santa Monica, CA: RAND Corporation, 2002.
2. Ursano RJ, Fullerton CS, Vance K, et al. Post-traumatic stress disorder and identification in disaster workers. *Am J Psychiatry* 1999; 156: 353–359.
3. Ursano RJ, Cerise FP, DeMartino R, et al. The impact of disasters and their aftermath on mental health. *J Clin Psychiatry* 2006; 67(1): 7–14.
4. Levin S, Herbert R, Szeinuk J, et al. Brief report: health effects of world trade center site workers. *Am J Ind Med* 2002; 42(6): 545–547.
5. Welch L, Ringen K, Bingham E, et al. Screening for beryllium disease among construction trade workers at Department of Energy nuclear sites. *Am J Ind Med* 2004; 46(3): 207–218.
6. Montiel-Lopez M, Corzo-Alvarez G, Chacin-Almarza B, et al. Prevelance and characterization of hearing loss in workers exposed to industrial noise of the turbogenerated electric plant of a petrochemical industry. *Invest Clin* 2006; 47(2): 117–131.
7. Nasterlack M, Hoffman G, Messerer P, et al. Epidemiological and clinical investigations among employee in a former herbicide production process. *Int Arch Occup Environ Health* 2007; 80(3): 234–238.
8. Robertson W, Robertson AS, Burge CB, et al. Clinical investigation of an outbreak of alveolitis and asthma in a car engine manufacturing plant. *Thorax*. 2007; 62(11): 981–990. Epub 2007 May 15.
9. Chen HI, Liou SH, Loh CH, et al. Bladder cancer screening and monitoring of 4,4'-methylenebis(2-chloroaniline) exposure among workers in Taiwan. *Urology* 2005; 66(2): 305–310.
10. OSHA (Occupational Safety and Health Administration). Medical Screening and Surveillance: Medical Screening, Available from: URL http://www.osha.gov/

SLTC/medicalsurveillance/screening.html [accessed 13 August 2007].

11. U.S. Department of Labor, Occupational Safety and Health Administration. Screening and Surveillance: A Guide to OSHA Standards. OSHA 3162, 2000; (reprinted).
12. Halperin WE, Ratcliffe J, Frazier TM, et al. Medical screening in the workplace: proposed principles. *J Occup Med* 1986; 28(8): 547–552.
13. Baker EL, Matte TP. Surveillance of occupational illness and injury. In: Halperin W, Baker EL, Mansoon RR, eds. *Public Health Surveillance.* New York: Van Nostrand Reinhold, 1992; 178–194, Chapter 13.
14. Koh D, AW TC. Surveillance in occupational health. *Occup Environ Med* 2003; 60(9): 705–710.
15. Lilis R, Miller A, Godbold J, et al. Radiographic abnormalities in asbestos insulators: effects of duration from onset of exposure and smoking: relationships of dyspnea with paranchymal and pleural fibrosis. *Am J Med* 1991; 20: 1–15.
16. Levin SM, Selikoff IJ. Radiological abnormalities and asbestos exposure among custodians of the New York City Board of Education. *Ann N Y Acad Sci* 1991; 643: 530–539.
17. NIOSH (National Institute for Occupational Safety and Health). Health Hazard Evaluations Program, 2007, Available from: URL http://www.cdc.gov/niosh/hhe/HHEprogram.html [Accessed 13 August 2007].
18. National Center for Health Statistics. *NHANES III Reference Manuels and Reports [CD-ROM].* Hyattsville, MD: Centers for Disease Control and Prevention, 1996.
19. Piccirillo JF, Merrit MG, Richards ML. Psychometric and clinimetric validity of the 20-Item Sino-Nasal Outcome Test (SNOT-20). *Otolaryngol Head Neck Surg* 2002; 126(1): 41–47.
20. Goldberg DP, Hillier VF. A scaled version of the General Health Questionnaire. *Psychol Med* 1979; 9(1): 139–145.
21. Blanchard EB, Jones-Alexander J, Buckley TC, Forneris CA. Psychometric properties of the PTSD Checklist (PCL). *Behav Res Ther* 1996; 34(8): 669–673.
22. Burney PG, Laitinen LA, Perdrizet S, et al. Validity and repeatability of the IUATLD (1984) Bronchial Symptoms Questionnaire: an international comparison. *Eur Respir J* 1989; 2(10): 940–945.
23. European Community Respiratory Health Survey. *Medicine and Health. EC Directorate General XIII. L-2920.* Luxembourg: Office for Official Publications, 1994.
24. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA* 1999; 282: 1737–1744.
25. Ewing JA. Detecting alcoholism. The CAGE questionnaire. *JAMA* 1984; 252: 1905–1907.
26. Leon AC, Olfson M, Portera L, et al. Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. *Int J Psychiatry Med* 1997; 27(2): 93–105.
27. Enright PL. Skloot G and Herbert 5. Standardization of Spirometry in Assessment of Responders Following Man-Made Disasters: The World Trade Center Worker and Volunteer Medical Screening Program. *Mt Sinai J Med 2008*; 75.
28. The New York City Department of Health and Hygiene. Clinical guidelines for adults exposed to the world trade center disaster. *City Health Inf* 2006; 25(7): 47–58.
29. Katz CL, Smith R, Holmes A. et al. A mental health program for ground zero rescue and recovery workers: cases and observations. *Psychiatry Serv* 2006; 57(9): 1335–1338.
30. de la Hoz RE, Shohet M, Chasan R, et al. Occupational toxicant inhalation injury: the World Trade Center (WTC) experience. *Int Arch Occup Environ Health* 2008; 81: 479–485.

Epidemiologic Research on Man-made Disasters: Strategies and Implications of Cohort Definition for World Trade Center Worker and Volunteer Surveillance Program

David A. Savitz, PhD, Rachael T. Oxman, BA, Kristina B. Metzger, PhD, MPH, Sylvan Wallenstein, PhD, Diane Stein, BS, Jacqueline M. Moline, MD, MSc, and Robin Herbert, MD

Department of Community and Preventive Medicine
Mount Sinai School of Medicine, New York, NY, USA

ABSTRACT

Studies of long-term health consequences of disasters face unique methodologic challenges. The authors focused on studies of the health of cleanup and recovery workers, who are often poorly enumerated at the outset and difficult to follow over time. Comparison of the experience at the World Trade Center disaster with 4 past incidents of chemical and radiation releases at Seveso, Italy; Bhopal, India; Chernobyl, Ukraine; and Three Mile Island, USA, provided useful contrasts. Each event had methodologic advantages and disadvantages that depended on the nature of the disaster and the availability of records on area residents, and the emergency-response and cleanup protocol. The World Trade Center Worker Monitoring Program has well-defined eligibility criteria but lacks information on the universe of eligible workers to characterize response proportions or the potential for distortion of reported health effects. Nonparticipation may result from lack of interest, lack of awareness of the program, availability of another source of medical care, medical conditions precluding participation, inability to take time off from work, moving out of the area, death, or shift from initially ineligible to eligible status. Some of these considerations suggest selective participation by the sickest individuals, whereas others favor participation by the healthiest. The greatest concern with the validity of inferences regarding elevated health risks relative to external populations is the potential for selective enrollment among those who are affected. If there were a large pool of nonparticipating workers and those who suffered ill health were most motivated to enroll, the rates of disease among participants would be substantially higher than among all those eligible for the program. Future disaster follow-up studies would benefit substantially by having access to accurate estimates of the number of workers and information on the individuals who contributed to the cleanup and recovery effort. ***Mt Sinai J Med 75:77–87, 2008.*** © 2008 *Mount Sinai School of Medicine*

Key Words: World Trade Center, man-made, disaster, cohort definition, surveillance.

Address Correspondence to:

David A. Savitz
Department of Community and
Preventive Medicine
Mount Sinai School of Medicine
New York, NY, USA
Email: david.savitz@mssm.edu

Epidemiologic investigations of the consequences of unexpected disasters, whether man-made or natural, are inherently limited by lack of anticipation and preparation for the research that follows. Although there are many studies of the health of those potentially affected by disasters, the unanticipated nature of the event makes it more difficult to carefully define the population of interest, measure exposure, and prepare for longitudinal health studies that follow. Studies of man-made and natural disasters, therefore, deviate from the ideal approach to study planning: specification of hypotheses in advance,

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20023

© 2008 Mount Sinai School of Medicine

identification of a suitable population for addressing the research question, accurate measurement of exposure, enrollment of participants to ensure comprehensive identification and a high rate of participation, administration of informed consent to participate in the study, and ensuring the ability to track participants' health status over time.

In the immediate aftermath of a disaster, long-term health concerns are of less concern than immediate threats to health and the environment. Gathering data that are needed to enumerate potentially exposed individuals is typically done post hoc by reconstructing information based on records of residences or jobs, anecdotal reports, or self-identification. Exposure measurements while the disaster is unfolding are rarely available, requiring reliance on historical exposure reconstruction through environmental modeling or individual reports of exposure. As concern shifts from acute to delayed effects of disaster, there is a need for accurate and comprehensive follow-up of exposed cohorts.

Accurately defining populations at risk of long-term health effects from the attacks on the World Trade Center (WTC) has been problematic from the outset, particularly with regard to those who assisted in the recovery efforts. There are several features of the disaster that pose challenges to accurately describing long-term health consequences of those experiences: the spectrum of exposure, including both environmental chemicals and psychological trauma; the location of the WTC in a densely populated residential and work area; and the many sources of workers who responded to help immediately and in the weeks and months that followed the attack.

In this discussion, we focus on the men and women who responded to the disaster and provided assistance after the WTC attacks, specifically on implications of cohort definition and recruitment for studies of exposure and health outcomes. First, we briefly review strategies employed to define exposed populations in past disasters, focusing on those performing rescue, recovery, restoration of services, and cleanup. Then, we consider, in some detail, the ideal and operational cohort definitions used in ongoing follow-up studies of WTC responders, considering the implications for inferences on exposure and health outcomes.

SUMMARY OF HEALTH RESEARCH ON PREVIOUS DISASTERS

Although each disaster situation is unique, prior disaster studies offer useful perspectives for evaluating the WTC worker health studies. Table 1 summarizes the key features of 4 broadly analogous disasters in which ionizing radiation or chemicals were released and affected large populations.[1–18] Like Bhopal, India, the WTC attack occurred in a highly urban setting, proximal to a large population of residents, students, and workers. As the US, is a more developed country, study of events here offers some advantages over studies of events in India and Ukraine such as an emergency response infrastructure and some records for enumerating residents and workers.

The environmental health concerns from visible dust exposure in the WTC attacks lend themselves to self-report in contrast to the ionizing radiation exposure at Chernobyl and Three Mile Island. However, ionizing radiation constitutes a specific agent with well-understood health effects, whereas the WTC exposures were an extremely complex mixture of agents contained in the dust from the collapsed buildings along with products of combustion. In addition, ionizing radiation is routinely monitored at sites of nuclear power plants, generating some baseline data and clues to an evolving disaster, whereas there was no corresponding information available for the WTC disaster.

The WTC attacks were immediately obvious, in contrast to the dioxin contamination at Seveso, which was not recognized for more than a week after it occurred. The intentional nature of the attack on the WTC tended to focus initial attention more on national security than on environmental health. Given that concern with security associated with the attacks on the WTC, the priority given to long-term health consequences was perhaps lower than in any of the other situations characterized in Table 1. In other disasters, the shift was from acute to chronic health effects, whereas after the WTC attacks, the shift was from acute health effects to the military and political response and then, after some delay, to chronic health concerns. National security concerns did not diminish in the aftermath of the disaster, reducing the attention given to the health of cleanup workers.

IDENTIFICATION OF RESPONDERS IN PREVIOUS DISASTER RESEARCH

Residents located near a disaster area can often be enumerated, even retrospectively, on the basis of existing records or by obtaining household rosters. Employees are typically amenable to identification

Table 1. *Summary of Disaster Experience and Epidemiologic Follow-up Studies for Long-term Health Effects: Seveso, Chernobyl, Three Mile Island, and Bhopal.*

Geographic location	Date of event	Primary exposure of concern	Residential population	Affected employed population	Emergency responder population	Cleanup worker population	Limitations and strengths
Seveso, Italy (Province of Milan)	July 10th, 1976	2,3,7,8–Tetrachloro-dibenzo-ρ-dioxin (TCDD) (dioxin)	Population: Highly documented. ~28,000 residents (1). 733 in most highly exposed residential area, Zone A (1,3). Method of determination: Based on mapping of TCDD soil concentrations (2) and records of residence. Medical/public health response: Surveillance measures were undertaken for 220,000 residents (1,3). Systematic medical testing and surveillance for randomly selected cohorts; 17,000 residents from Zones A, B and R received systematic medical testing through 1984 (1,5). All school children (32,200) received skin examinations (1,4). Health outcomes tracked through records of health care providers and vital statistics.	Population: Highly documented. Industrie Chihiche Meda Societa (ICMESA) chemical plant workers Method of determination: Chemical plant documents. Medical/Public health response: Workers were evaluated in specialized health studies as one of the subgroups in a randomly selected cohort.	Population: Not documented. Local emergency response Medical/public health response: Emergency responders did not receive special focus in heath studies.	Population: Highly documented. Decontamination workers. Method of determination: Detailed records of "cleanup" employees including medical entrance exam. Medical/Public health response: Workers were examined before entrance into Zone A and again after 9 mo of decontamination work (4) to ensure limited exposure.	Limitations: Presence of TCDD[a] was unknown until 9–10 days after the incident (2–4). Strengths: Residence registrars made identification of residents in severely affected areas accurate and nearly complete.

(continued overleaf)

Table 1. *(Continued).*

Geographic location	Date of event	Primary exposure of concern	Residential population	Affected employed population	Emergency responder population	Cleanup worker population	Limitations and strengths
Three Mile Island Nuclear Power Plant (10 miles away from Harrisburg, PA)	March 28th, 1978	Ionizing radiation (total release of ~2.5–10 million curies, radioactivity)	Population: Highly documented. ~35,900 residents living within a 5-mile radius of the Three Mile Island (TMI) Nuclear Plant (11). Method of Determination: Based on records of residence.	Population: Highly documented. 533 TMI Nuclear Plant workers (12). Method of determination: Based on plant employment records.	Population: None. No emergency response–no explosion or fire occurred.	Population: Highly documented. >1,000 (13). Method of determination: detailed records of "cleanup" employees.	Strengths: Affected population geographically contained. Accurate and up-to-date residence information. Complete employee roster for the TMI Plant.
			Medical/Public health response: TMI Population Registry. Door-to-door canvassing of 5-mile radius around TMI to administer standardized questionnaires (11).	Medical/Public health response: Permanent TMI Employees as of March 1st, 1979 were surveyed by telephone interview (12).		Medical/Public health response: Worker exposure monitored very closely. As of 1990 average worker dose 62.17 person–Sieverts (Sv) (14)	Able to utilize effective state and national organizations with experience in registering communities. Also had sufficient resources for follow-up and data processing.
Bhopal, India (population of 800,000–1,000,000) (15,16)	December 2nd to 3rd, 1984	Methyl isocyanate (MIC) (27–40 tons (16–18)) and decomposition by-product hydrogen cyanide.	Population: Poorly documented. 520,000 exposed residents; 200,000 residents lived in high exposure areas (15,16). Method of determination: Areas of exposure defined as severe, moderate and mild according to mortality of both humans and animals (15).	Population: Not documented. Known to be Night Union Carbide Plant Workers.	Population: Not documented.	Population: No substantial cleanup effort and no employee roster. Medical/Public health response: No follow-up.	Limitations: Poorly defined population. Exact number of dead is unknown because official records only documented those who died in hospitals and bodies were burned soon after.

			Medical/public health response: Bhopal Gas Disaster Research Center in 1985, registered ~80,000 residents from severely, moderately and mildly affected areas as well as ~15,000 from unaffected Bhopal areas (15–17). Population Based Cancer Registry in Bhopal in 1986.				
Chernobyl town and neighboring town, Pripyat.	April 26th to 27th, 1986	Ionizing radiation (mostly in Belarus, Russian Federation, and Ukraine. Dispersed across Europe)	Population: Partially documented. ~340,000 evacuated from area (6,7) ~120,000–135,000 residents living within 30-km radius of the Power Plant (6,8) ~2,326,000 exposed residents, evacuees, re-settlers and children born to exposed parents living in the exposed areas of Russia, Ukraine and Belarus. Method of determination: Based on proximity to disaster site, dates of residence, self report, etc. Medical/public health response: 50,000 radiological examinations on the evacuees, residents and workers.	Population: Partially documented. Chernobyl Nuclear Power Plant workers. Method of determination: ARS diagnosis in aftermath of disaster. Medical/public health response: Only those Nuclear Plant workers diagnosed with ARS received specialized follow-up care and had their medical outcomes tracked.	Population: Partially documented. ~600 emergency personnel participated in fire department and helicopter response (7). Method of determination: ARS diagnosis in aftermath of disaster. Employment records. Medical/public health response: All surviving individuals of the initial 237 suspected cases of acute radiation sickness have undergone long-term monitoring (7,9).	Population: Partially documented. ~300,000–600,000 (7,8) Method of determination: 537,000 liquidators registered in the Chernobyl registries of Russia, Ukraine and Belarus (9). ~600,000 Liquidators were issued certificates of service onsite. Once registered, participants were required to undergo annual exams (7,10). Russia, Belarus and Ukraine all have established clinics for the monitoring of liquidators as well as institutions dedicated to providing them with specialized medical care and rehabilitation (9).	Limitations: No complete roster of Chernobyl liquidators. Reliant on individual's account of duties and location in estimating radiation dosage. Large incentive to forge liquidator status as it was linked to many social and medical benefits. Liquidators were not geographically contained. Strengths: Great incentive for liquidators to enroll in national registries so as to gain access to benefits.

---

[a] 2,3,7,8 - Tetrachlorodibenzo-ρ-dioxin

and enumeration through payroll lists or time sheets for employees who were present at work on a specific day. In sharp contrast, responders are comprised of an ad hoc assembly of individuals brought in on an emergency basis, often in an unplanned, poorly documented manner. In the initial phase of response, help is often sought from wherever it can be found, including designated emergency response workers, local residents and workers, and volunteers seeking to be of assistance. They are identified solely by their activity of participation in the cleanup operation, with no counterpart to residential or employment records to reconstruct the roster or source population from which they came.

Because the Seveso dioxin contamination episode was not identified until the chemical release was over and health risks were known to be present,[2–4] cleanup workers were engaged in a coordinated, well-documented manner; medical examinations were made before beginning work and again after 9 months of effort.[4] Similarly, in the Three Mile Island nuclear accident, around 1000 cleanup workers were engaged and individually monitored for ionizing radiation exposure by using personal dosimetry.[13,14] In Bhopal, there has been little documentation of response activities or the number or source of workers involved.

The Chernobyl nuclear disaster offers the most relevant analogy to the WTC worker health studies. The accident at Chernobyl in 1986 engaged emergency response workers during the days after the accident, with levels of ionizing radiation producing acute radiation sickness symptoms in approximately 237 patients, of which 145 were diagnosed with acute radiation syndrome (ARS).[7] Later, liquidators were recruited from Belarus, Russia, and Ukraine to come to the accident site and decontaminate the facility. Around a half million of these liquidators were engaged to decommission the plant, drawing on populations of 3 different countries.[9] (Table 1). The scale of the operation resulted in a system of clinics and health data systems focused on such workers, and a sizable volume of sophisticated epidemiologic studies have resulted. The systematic approach to hiring and monitoring these workers supports epidemiologic follow-up studies. The WTC rescue and recovery workers were not hired or documented in a comparable manner.

WORLD TRADE CENTER WORKER MONITORING PROGRAM: ELIGIBILITY AND ENROLLMENT

The ongoing World Trade Center Worker and Volunteer Medical Screening and Medical Monitoring Program, referred to here as the WTC Worker Monitoring Program, was established in July 2002 to provide comprehensive medical monitoring examinations for workers and volunteers who were engaged in rescue, recovery, restoration of services, and cleanup after the attacks on the WTC. A separate monitoring program was established by the Fire Department of New York for New York City firefighter responders. A federally funded treatment program for both cohorts was added in November 2006.

The WTC Worker Monitoring Program includes a clear operational definition of eligibility based on an exposure matrix of tasks performed combined with time and duration of work at the site. Initially, eligibility was narrowly defined to include only those individuals who had worked within the geographic area of "the pile" and its perimeter. Time requirements consisted of working 24 hours during the first 3 days following the attacks or a total of 10 days at the site in the month of September 2001.

Over time, as clinic capacity increased, eligibility criteria underwent multiple expansions. The task requirement stayed the same, with workers or volunteers required to have performed rescue, recovery, cleanup, or restoration of essential services. The geographic scope expanded, however, to include Manhattan below Canal Street, the barges, the Staten Island landfill, and several other locations where work did not necessarily occur within the stated geographic boundaries (eg, vehicle maintenance on WTC-dust-contaminated trucks that were transported to garages in Brooklyn and Queens). The time requirements were expanded to include several alternatives: 4 hours between September 11 and 15, 2001; 24 hours between September 11 and 30, 2001; or 80 hours from September 11 through December 30, 2001, which is the date that the fires at the site finally ended. Ascertainment of eligibility is conducted by staff at a telephone bank and is based on self-reported exposures.

Eligibility criteria for enrollment also changed in response to availability of alternate programs for other categories of workers. For example, federal and New York state workers were initially *not* eligible for the WTC Worker Monitoring Program, but as monies for programs for these workers ran out, the eligibility criteria were changed to allow those individuals to enroll.

Several groups were eligible for inclusion in the program when they met special requirements. Port Authority Trans–Hudson train (PATH) workers were eligible if they met separate time and location requirements, working in the PATH tunnels for at least 24 hours during the period from February to July of 2002. This exception was made because cleanup of the PATH tunnels, which were heavily damaged and contaminated by a significant amount of dust and debris, did not begin until February 2002. Employees of the Office of the Chief Medical Examiner were eligible despite the amount of time worked or the location of their work.

Given the nature of cleanup operations and application of eligibility criteria based on engaging in *activities* rather than satisfying specific *administrative criteria* (eg, belonging to specific worker groups), there is no roster of those individuals eligible to participate. That is, there is no listing of all individuals who were engaged in specific activities at locations and in time periods of interest that would qualify them for participation in the WTC Worker Monitoring Program. We have a roster of individuals deemed eligible by the telephone bank who chose not to make at least 1 visit to the clinical center, and a roster of those who had at least 1 visit to the clinical center. We have essentially no information on the number or identity of those who did not contact the telephone bank.

Extensive efforts were made to make diverse, eligible individuals aware of the program through newsletters and mailings and widespread media coverage, especially in the New York area. Telephone calls have been received from 29,591 presumably eligible individuals based on the eligibility criteria in effect at the time of their call or who were initially not eligible but became eligible as the criteria were expanded. It is inevitable, of course, that some of those who are eligible have not participated even at the level of a telephone call, which raises concern about their potential need for expert clinical care. However, our focus is on the potential for incomplete coverage of those eligible to affect research on the relation between WTC exposures and health outcomes. The loss of otherwise eligible participants, even if truly random, incurs some loss of statistical power because of the smaller number of responders available to study. Moreover, there is a real possibility for such losses to be nonrandom with regard to the exposures and outcomes of interest. That is, when we evaluate the relation between WTC exposures and long-term health consequences by studying the individuals who participated in the WTC Worker Monitoring Program, we are susceptible to drawing different inferences than we would have if all those who were eligible had participated in the program.

Initial estimates of the number of workers eligible for the program by Mount Sinai staff was approximately 40,000, subject to some uncertainty. This number was derived from estimates given by several major groups involved in the recovery effort, including the American Red Cross and major labor unions. The estimate excluded employees of the Fire Department of New York, which had their own program, and both federal employees and New York State employees who were not eligible for the Mount Sinai program because they were able to participate in employer-coordinated programs. The New York City Department of Health and Mental Hygiene, which is conducting their own health assessments of the population exposed to the World Trade Center disaster through the World Trade Center Health Registry,[19,20] has published estimates that 91,469 workers and volunteers would be eligible for their registry. However, eligibility criteria for their registry were notably different from the World Trade Center Worker Monitoring Program: 1) The time requirement for participation in the World Trade Center Registry was "1 shift" between September 11, 2001, through June 30, 2002, which is far less restrictive than the time period required for participation in the World Trade Center Worker Monitoring Program; 2) Their estimate includes 26,480 volunteers, among whom some unknown fraction would not meet the eligibility criteria for our population; 3) Their estimate includes 8897 workers from State Agencies as well as 5122 Federal employees who were initially not eligible for participation but who were included later; and 4) They include 13,500 firefighters who have their own program and were, thus, not part of our monitoring. Some (but not all) of these considerations could be addressed to reconcile the various estimates and to generate information on the approximate response proportion among those eligible for the World Trade Center Worker Monitoring Program. In the absence of clear documentation, the number of workers meeting specific eligibility criteria is subject to substantial uncertainty. With nearly 30,000 eligible individuals having made at least telephone contact with the World Trade Center Worker Monitoring Program, we estimate that the response proportion is currently between 50% and 70% and increasing steadily as new participants continue to call.

Given the broad eligibility criteria for participation in the WTC Worker Monitoring Program based on the location, time, and duration of activities, the cohort includes workers with wide variation in type and magnitude of exposure. Some responders experienced intense, but brief, exposure during the first days after the disaster with intense exposure to contaminated air and unstable physical working conditions, as well as psychological trauma resulting from recovery of body parts and uncertainty surrounding what was happening. Workers who started working at a later date, but worked for a longer time, had a different exposure experience both physically and psychologically.

IMPLICATIONS OF INCOMPLETE COVERAGE

To empirically describe the nature of incomplete enrollment and its implications for assessing the impact of WTC exposures on worker health, we would need data on those persons who were eligible but did not enroll. As indicated above, this information is not available, leaving us only with informed speculation based on the general experience in other studies and integrated with the specific features of the WTC Worker Monitoring Program.

The reasons that some members of the universe of eligible workers have not participated are likely to be diverse. We consider a spectrum of possible pathways by which this could have occurred and the implications of each with regard to both exposure and health status. To the extent that such losses are nonrandom with regard to exposure, inferences of possible health risk may be exaggerated or understated. If more highly exposed workers participated in the monitoring program and suffered increased health risks as a result of their elevated exposure, we would overstate health risks if their experience is assumed to apply to all those eligible for the WTC Worker Monitoring Program and, conversely, if the less exposed workers participated preferentially. To the extent that participation is related to health status, we will overstate the effect of their experience if the sickest workers are most likely to participate and understate the effect of their experience if the healthiest workers are preferentially enrolled.

However, neither of those scenarios necessarily results in a distortion of the measured relation between exposure and disease.[21] The opportunity to compare health outcomes among participants with greater and lesser degrees of exposure is advantageous in that the unknown losses of eligibles are less likely to be problematic for comparisons among participants than for comparison of participants to external populations. Only a situation in which nonparticipation results in selective inclusion (or exclusion) *jointly by exposure and health status* will distortion of exposure–response result. If those workers who are least exposed *and* healthiest fail to participate, or those who are most exposed *and* sickest fail to participate, we will underestimate the magnitude of exposure impact on health. Similarly, if those who are least exposed *and* sickest fail to participate, or those who are most exposed *and* healthiest fail to participate, we will overestimate the magnitude of exposure's impact on health. Issues of joint determination of participation based on exposure and health outcomes are more complex and less plausible than an influence of exposure or health status alone.

POSSIBLE REASONS FOR INCOMPLETE COVERAGE

No Interest in the Program

As in most voluntary activities, a proportion of potential participants will choose, often with relatively modest deliberation, to decline to enroll simply for lack of time or interest. Although the potential health consequences of WTC involvement are very significant to many of those exposed, a subset may either prefer not to dwell on their past experience or find medical-monitoring activity irrelevant in their day-to-day lives because of the absence of known health problems. Despite the magnitude of the event, some fraction of workers, likely among the healthiest, would assign medical monitoring to a low priority.

No Awareness of Program or Eligibility

Given the diversity of groups involved in the cleanup effort, it remains quite possible that some of the individuals who participated and would qualify for the WTC Worker Monitoring Program are simply unaware of the program's existence or their eligibility to obtain clinical care through the program. Those most likely not to be reached are those who are no longer affiliated with their former organization (eg, union, employer) and not attending to media. Both leaving employment and social isolation would likely be associated with worse health status, on average. Those with the most immediate and intense exposure in the cleanup effort are more likely to be aware of their eligibility than those with delayed or less intense exposure, eg, on the waste pile rather than at the building site.

Medical Care Outside of the Program

Those individuals who have ongoing medical care and are pleased with their current health care provider may believe that there is no benefit from clinical services provided by the WTC Worker Monitoring Program. Those who have private medical care (and are, thus, of greater economic means) may be less likely to perceive benefit from the program. Alternatively, those who are receiving ongoing care for a serious illness may be reluctant to obtain additional medical services. Greater economic means would be predictive of more favorable health status, whereas ongoing care for a chronic illness would be indicative of less favorable health status.

Illness

Workers who are hospitalized or severely debilitated may find it unfeasible to come to the clinic site for logistical reasons. Although efforts have been made to accommodate all who are interested in being seen, doing so does require a minimum degree of mobility and function. Obviously, those who are physically unable to come to the clinic sites would have much worse health status than those who are able to do so.

Time Off

Given the need to travel to the clinic and be absent from work during normal work hours, those whose jobs do not facilitate clinic participation would be less likely to do so. Among the spectrum of responders, this would encourage selective enrollment of those relatively few whose employers provided paid time off for examinations. Most have had to use nonwork time to take advantage of clinic services.

Moved

While there is a national network available to those individuals who are eligible but do not live in WTC area, this would be less well known and perhaps seen as less salient among those outside the area. Those residing outside the New York/New Jersey metropolitan area tend to fall into 4 categories: 1) urban search and rescue teams and other groups of responders who had lived outside the area at the time of the disaster; 2) those who retired at the usual age then moved away; 3) those who retired early or were disabled because of WTC-related illnesses and who relocated for economic and/or health reasons; and 4) those who relocated for economic opportunity or other reasons. Obviously, these very different reasons for residing outside the area have different implications for health, with illness-related retirement, of course, being associated with poor health status and moving for economic reasons being a more favorable sign with regard to health.

Deceased

The numbers of individuals who have died would likely be small, whether related or unrelated to exposures received in the WTC cleanup and recovery activities, given the selectively healthy population of volunteers at the site and the relatively brief period of time since September 11, 2001. Such losses would deplete the cohort of less healthy individuals compared with those who are still living.

Initially Ineligible but Became Eligible Through Changes in Criteria

Those workers who were initially ineligible because of duration or location of work or for administrative reasons because their employer may be less likely to have learned of their eligibility compared with those who were aware at the beginning.

IMPLICATIONS OF INCOMPLETE ENROLLMENT ON INFERENCES REGARDING EXPOSURE AND HEALTH OUTCOMES

There are multiple plausible reasons that eligible workers may not enroll in the WTC Worker Monitoring Program, as enumerated above. The greater the proportion of eligibles who enrolled, the lower the risk of selection bias resulting in distortion of disease rates and measures of association between exposure and disease. Acknowledging that participation is less than complete because of unknown (and probably unknowable) extent, the next concern is whether those who have enrolled are selectively healthy or ill compared with those who did not enroll. There is much interest in comparing the rates of illness among WTC Worker Monitoring Program participants with the general population, as determined through disease registries, health surveys, and other means. Where rates of disease are notably higher in WTC responders, the inference is that their work experience is likely to have contributed to that elevation in risk. To the extent that the rate of disease among participants is markedly different from the rate of disease that would have been observed among all those who actually responded (eligibles), inferences will be inaccurate.

Although we are lacking data to directly address the magnitude or direction of distortion comparing those enrolled with the entirety of those who were eligible, there is a reasonable basis for speculating on tendencies for participants to be selectively healthy or unhealthy. In general, the presence of illness makes any health or medical activity (clinical or research) more salient and, thus, more worthy of investing one's time. Other considerations being equal, those who have health concerns, symptoms, or illnesses are more likely to take advantage of the WTC Worker Monitoring program than those who are in good health for whom health is simply not a pressing concern. The greatest risk of having spuriously elevated rates of disease among WTC Worker Monitoring Program participants is the potential for enrollees to represent the sickest among a much larger pool of eligible workers. The magnitude of this distortion depends on the proportion of eligibles who participated, the magnitude of the tendency for the least healthy to be most likely to enroll, and the impact of compensating tendencies in the other direction, favoring enrollment by the healthiest individuals.

There are several reasons to expect competing tendencies that tend toward selective participation by the healthiest responders. Lack of awareness of the program and inability to take time from work would result in nonparticipation, and both are likely to be associated with lower health status, on average. Obviously, loss of those who are too severely ill to participate or those who are deceased, and those who already have medical care because of serious illness and, thus, do not participate would result in a more favorable health profile among participants. The relative magnitude of these competing tendencies is unknown, and, thus, the net impact of nonresponse on estimated disease rates cannot be estimated with certainty.

Beyond the interest in characterizing the health experience of workers in aggregate, we are interested in gradients in risk of disease associated with *level* of exposure among participants in the monitoring program. That is, we will compare health outcomes among those with greater exposure to those with lesser exposure. In such comparisons, the potential for erroneous inference is much less because all workers (both those with higher and lower exposure) are subject to selective influences for participation. To the extent that meaningful gradients in exposure can be identified among the participants, requiring separation of those with little or no exposure from those with more substantial exposure, the strength of inferences will be markedly enhanced. For dose–response gradients to be influenced by selective participation, there would have to be more complex patterns of response jointly by both exposure and disease status.

In the face of these competing influences of unknown magnitude, sensitivity analyses would be of value to help characterize the certainty of findings linking WTC exposures and disease. Such an approach would entail various assumptions about the likely magnitude of each of the many pathways of loss of eligible participants and incorporating those probabilities into the distribution of possible values for the measure of association.[22] Limited data on each of the mechanisms of nonparticipation would require speculation and generate a wide range of possibilities, but they would help to bound the potential biases. A formal sensitivity analysis is beyond the scope of this more general discussion of the concerns but warranted as a means of evaluating the validity of specific findings as they emerge.

LESSONS FOR DISASTER EPIDEMIOLOGY

Although evaluation of health consequences of disasters must, by definition, be unplanned because of the unpredictable nature of the precipitating event, some suggestions can be offered for improving inferences regarding long-term health effects without distracting from the need for immediate action. Key to all future studies is the need for some information on the size of the workforce that participated, ideally including their demographic profile or, even more optimally, their registration for future follow-up. Even if such information is unattainable in the immediate hours and days after the event, in the weeks that follow, an increasingly organized and fully documented approach should become more feasible.

The evolution from acute health concerns to chronic health concerns can be anticipated. Provision of services for acute, physical trauma inevitably extends to long-term physical and mental health concerns. Perhaps some of the long-term consequences could be averted by more complete follow-up and care from the beginning. More systematic enrollment and monitoring would not only provide a firmer basis for inferences regarding health risks but would also result in better clinical care. Such plans should be developed as part of a comprehensive disaster management strategy, building upon and not detracting in any way from the acute health concerns associated with such disasters.

ACKNOWLEDGMENTS

This work was supported by the Centers for Disease Control and Prevention and the National Institute for Occupational Safety and Health ([R.H.] grant 5U1O 0H008232 and grant 3U10 OH008225).

REFERENCES

1. Pocchiari F, Silano V, Zampieri A. Human health effects from accidental release of tetrachlorodibenzo-*ρ*-dioxin (TCDD) at Seveso, Italy. *Ann N Y Acad Sci* 1979; 320: 311–320.
2. Bertazzi PA, Bernucci I, Brambilla G, et al. The Seveso studies on early and long-term effects of dioxin exposure: A review. *Environ Health Perspect* 1998; 106(suppl. 2): 625–633.
3. Pocchiari F, Silano, Zapponi G. The chemical risk management process in Italy. A Case Study: the seveso incident. *Sci Total Environ* 1986; 51: 227–235.
4. Bertazzi P, di Domenico A. Health consequences of the seveso, Italy accident. In: Schecter A, Gasiewicz TA. eds. *Dioxins and Health*. New York: John Wiley and Sons, 827–853.
5. Brambilla P, Gerthoux P, Beretta C, et al. Seveso: biological monitoring of dioxin pollution. An example of environmental medicine. Multicentric and interdisciplinary applied research. *G Ital Med Lav Ergon* 1998; 20(3): 197–199.
6. Mould RF. *Chernobyl Record: The Definitive History of the Chernobyl Catastrophe*. Bristol, CT: Institute of Physics Publishing, 2000.
7. UNSCEAR. Annex J: Exposures and effects of the Chernobyl accident, Report to the United Nations General Assembly June 2000.
8. Baverstock K, Williams D. The chernobyl accident 20 years on: an assessment of the health consequences and the international response. *Environ Health Perspect* 2006; 114(9): 1312–1317.
9. Bennett B, Repacholi M, Carr Z, eds. *Health Effects of the Chernobyl Accident and Special Health Care Programmes*. Geneva: WHO Report of the UN Chernobyl Forum Expert Group "Health", 2006.
10. Pitkevitch VA, Ivanov VK, Tsyb AF, et al. Exposure levels for persons involved in recovery operations after the Chernobyl accident. *Radiat Environ Biophys* 1997; 36: 149–160.
11. Goldhaber MK, Tokuhata GK, Digon E, et al. The three mile island registry. *Public Health Rep* 1983; 98(6): 603–609.
12. Kasl SV, Chisholm RF, Eskenazi B. The impact of the accident at the three mile Island on the behavior and well-being of nuclear workers. *Am J Public Health* 1981; 71(5): 472–482.
13. World Nuclear Association. Briefing Paper #48. Three Mile Island, London, 2001.
14. Miller KL. Radiologic history exhibit: the nuclear reactor accident at three mile Island. *Radiographics* 1994; 14: 215–224.
15. Sundar Rao PSS. Statistical issues in assessing the toxic affects of bhopal gas disaster. *Int Stat Rev* 1994; 61(2): 223–229.
16. Dhara RV. Dhara R. The union carbide disaster in Bhopal: a review of health effects. *Arch Environ Health* 2002; 57(5): 391–404.
17. Broughton E. The Bhopal disaster and its aftermath: a review. *Environ Health* 2005; 4(1): 1–6.
18. Dhara RV, Dhara R, Acquilla SD, Cullinan P. Personal exposures and long-term health effects in survivors of the union carbide disaster in Bhopal. *Environ Health Perspect* 2002; 110(5): 487–500.
19. Brackbill RM, Thorpe LE, DiGrande L, et al. Surveillance for World Trade Center disaster health effects among survivors of collapsed and damaged buildings. *MMWR Surveill Summ* 2006; 55: 1–18.
20. Murphy J, Brackbill RM, Thalji L, et al. Measuring and maximizing coverage in the World Trade Center Health Registry. *Stat Med* 2007; 26: 1688–1701.
21. Rothman KJ, Greenland S, eds. *Modern Epidemiology*. 2nd ed. Philadelphia, PA: Lippincott, Williams, & Wilkins, 1998.
22. Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology* 2003; 14: 451–458.

Mount Sinai Journal of Medicine 75:89–100, 2008

Upper and Lower Respiratory Diseases After Occupational and Environmental Disasters

David J. Prezant, MD,[1,2,3] Stephen Levin, MD,[4] Kerry J. Kelly, MD,[2,5] and Thomas K. Aldrich, MD[3]

[1] Office of Medical Affairs, Fire Department of the City of New York (FDNY), New York, NY, USA
[2] FDNY World Trade Center Monitoring and Treatment Program, New York, NY, USA
[3] Pulmonary Division, Albert Einstein College of Medicine and Montefiore Medical Center, New York, NY, USA
[4] Irving Selikoff Occupational Medicine Clinic, Mount Sinai School of Medicine, New York, NY, USA
[5] Bureau of Health Services, Fire Department of the City of New York (FDNY), New York, NY, USA

ABSTRACT

Respiratory consequences from occupational and environmental disasters are the result of inhalation exposures to chemicals, particulate matter (dusts and fibers) and/or the incomplete products of combustion that are often liberated during disasters such as fires, building collapses, explosions and volcanoes. Unfortunately, experience has shown that environmental controls and effective respiratory protection are often unavailable during the first days to week after a large-scale disaster.

The English literature was reviewed using the key words–disaster and any of the following: respiratory disease, pulmonary, asthma, bronchitis, sinusitis, pulmonary fibrosis, or sarcoidosis.

Respiratory health consequences after aerosolized exposures to high-concentrations of particulates and chemicals can be grouped into 4 major caterogies: 1) upper respiratory disease (chronic rhinosinusitis and reactive upper airways dysfunction syndrome), 2) lower respiratory diseases (reactive [lower] airways dysfunction syndrome, irritant-induced asthma, and chronic obstructive airways diseases), 3) parenchymal or interstitial lung diseases (sarcoidosis, pulmonary fibrosis, and bronchiolitis obliterans, and 4) cancers of the lung and pleura.

This review describes several respiratory consequences of occupational and environmental disasters and uses the World Trade Center disaster to illustrate in detail the consequences of chronic upper and lower respiratory inflammation. ***Mt Sinai J Med 75:89–100, 2008.*** © *2008 Mount Sinai School of Medicine*

Key Words: World Trade Center, Disaster Medicine, Occupational Medicine, Respiratory Disease, Asthma, Chronic Rhinosinusitis, Gastroesophageal Reflux.

Address Correspondence to:

David J. Prezant
Office of Medical Affairs
Fire Department of the City of
New York (FDNY)
FDNY World Trade Center
Monitoring and Treatment
Program
Brooklyn, New York, USA
Email: prezand@fdny.nyc.gov

Nearly all types of environmental disasters (ie, fire, building collapse, explosion, volcano, earthquake, hurricane, flood, etc.) liberate high concentrations of aerosolized particulate matter. In addition to the acute and chronic inflammatory effects of particulate matter, these exposures are complicated by simultaneous exposures to the incomplete products of combustion that are liberated during disasters such as fires, building collapses, explosions, and volcanoes and inhaled either as fumes, vapors and gases, or as coated particulates. Unfortunately, experience has shown that environmental controls and effective respiratory protection (eg, fit-tested P-100 respirator) are often unavailable during the first days to a week after a large-scale disaster,[1] and adherence to proper use guidelines may be suboptimal in the weeks and months that follow.[1]

Respiratory health consequences after aerosolized exposures to high concentrations of particulates and chemicals during a disaster can be

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20028

© 2008 Mount SinaiSchool of Medicine

grouped into 4 major categories. First, *upper respiratory disease*, usually but not always manifests as cough and has been caused by reactive upper airway dysfunction syndrome (RUDS), a chronic inflammation of nose and sinuses. Second, *lower respiratory disease*, also most commonly manifests as cough but often has associated chest tightness, shortness of breath, and exercise intolerance, and has been caused by reactive (lower) airway dysfunction syndrome (RADS) and irritant-induced asthma, types of asthmatic bronchitis that may lead to chronic obstructive airways diseases. Third, *parenchymal or interstitial lung diseases* occur such as sarcoidosis, pulmonary fibrosis, bronchiolitis obliterans (fixed airway obstruction) and incidental pulmonary nodules. Fourth, *cancers of the lung and pleura* manifest.

Gastroesophageal reflux is an important factor in upper and lower respiratory diseases such as rhinosinusitis, laryngitis, tracheobronchitis, and asthma. It is unclear whether this reflux is a cause, effect or complication, but it is clear that concurrent treatment is important if therapeutic success is to be optimized. For these reasons, some investigators prefer to describe this group of upper and lower respiratory diseases as aerodigestive diseases. Disaster-related gastroesophageal reflux disease (GERD) is a new finding after the World Trade Center (WTC) attack. We could find no reports after other major disasters.

The epidemiology, pathophysiology, and clinical features of occupational and environmental respiratory diseases have been the subject of numerous review articles and monographs. For example, there is a growing body of scientific literature that demonstrates associations between chronic exposure to air pollution and the increased prevalence or severity of asthma,[2] chronic obstructive pulmonary disease,[3] cancer,[4] and ischemic heart disease.[5] Occupational asthma and chronic obstructive pulmonary disease can occur after chronic exposure to low and high molecular-weight compounds.[6,7] Airway hyper-reactivity has been shown after an acute mixed exposure to particulates and chemicals.[8] In miners[9] and silicon[10] and construction workers,[11] bronchial hyper-reactivity was documented several years after chronic low-level exposures to airborne particulates. Interstitial lung diseases and pulmonary fibrosis after exposure to inorganic dust and fibers, commonly referred to as pneumoconioses, have been well described after chronic occupational exposure to asbestos,[12] coal,[13] silica,[14] and hard metals such as cobalt.[15] Pulmonary fibrosis can also result from prolonged hypersensitivity pneumonitis as a result of environmental or occupational exposure to molds, animal proteins, and other allergens.[16,17] Sarcoidosis or sarcoid-like granulomatous disease has been described after exposure to silica, calcium carbonate, glass fibers, asbestos, hydrocarbons, wood burning, and metal dusts or fumes including beryllium, cobalt, copper, gold, aluminum and zirconium.[18–20] And, of course, the epidemiologic links between environmental and occupational exposure and lung cancer and malignant mesothelioma have been described in numerous articles.[12,21]

In contrast, there is a paucity of published data on respiratory effects after acute or subacute exposures to respirable particulates and chemical vapor, gases or fumes during man-made or natural disasters. After smoke inhalation, bronchial hyperreactivity may occur within hours,[22–24] and 1 study showed persistence in 11 of 13 subjects at 3-month postexposure.[24] Following the Mt. St. Helens eruption in 1980, hospital visits for pediatric asthma increased in Seattle, Washington, presumably related to exposures to respirable acidic volcanic dust.[25] Building materials (eg, gypsum-containing wallboard, cement, glass, and man-made vitreous fibers, heavy metals), aerosolized in collapse zones after fires, earthquakes, and other similar disasters, have been implicated in inflammatory syndromes that involve the respiratory mucosa and lung parenchyma.[26,27] Disasters resulting in acute exposures to chemical gases and vapors can result in airway hyperreactivity and clinical asthma, and when this occurs in nonsmokers without a prior history of asthma or allergies, it is often referred to as reactive airway dysfunction syndrome, a variant of irritant-induced asthma.[7] After the explosion of the Union Carbide Chemical Plant in Bhopal, India, studies documented an increased loss of pulmonary function compared with normal aging and increased presence and severity of obstructive airway disease.[28–30] Eosinophilic pneumonitis has been reported in US military personnel deployed in or near Iraq; all were smokers and most (78%) were new smokers with acute to subacute exposure to fine airborne sand or dust. Many required mechanical ventilation and corticosteroid treatment, and all survivors had normal or near normal spirometric measurements.[31]

Recently, a great deal of attention has been focused on respiratory health consequences of WTC dust exposure. This article concentrates on the WTC disaster because it is the largest urban disaster in modern time. Much has been and will be learned concerning its impact on the respiratory health of rescue workers, volunteers, and residents, and this knowledge should be transferable to any future disaster. Although case reports and small studies exist that describe upper and lower airway disease in disasters defect the WTC attack, none have been as well studied, and therefore, an epidemiologic comparison in terms of disease incidence rates is not possible except to note that thousands of WTC rescue and recovery workers have enrolled in federally funded treatment programs.

The September 11th, 2001, terrorist attacks on the WTC led to the fiery collapse of the Twin Towers and WTC Building 7. Approximately, 70% of the buildings' structural components were pulverized,[32] and the collapse produced a plume of dust and ash that spread throughout lower Manhattan and beyond. Furthermore, combustion of approximately 91,000 L of aircraft fuel ignited numerous structural fires, many of which smoldered until mid-December of 2001. Contaminants such as asbestos, hydrochloric acid, polychlorinated biphenyls (PCBs), silica, and heavy metals were found in the dust and ash from the WTC collapse.[32,33] The fires at the site created toxic combustion products, such as polycyclic aromatic hydrocarbons (PAHs), dioxins, volatile organic compounds, and various other known carcinogenic compounds.[32,33]

There is abundant evidence that the upper and lower respiratory symptoms experienced by WTC rescue and recovery workers were due to the dust they inhaled and ingested. A clear exposure–response gradient has been demonstrated, with the highest symptom prevalence found in those directly exposed to the dust cloud, arriving during the morning of September 11th, 2001.[34,35] Ninety-five percent of the respirable WTC dust was composed of large particulate ($\geqslant 10\ \mu$m in diameter) matter.[32] Particles of this size have conventionally been thought to be filtered by the upper respiratory tract, rarely entering lower respiratory structures.[36] However, there are a number of reasons to expect lower airways also to be at risk from the dust cloud. First, it has been shown that alkaline dust impairs nasal clearance mechanisms,[36] and most WTC dust samples had a pH greater than 10.[32] Second, the nasal filtration system is optimally functional during restful breathing. However, WTC rescue and recovery workers, as a consequence of their work activities (moderate to high level physical exertion), were breathing at high-minute ventilations where mouth breathing predominates. Third, although only 5% of the WTC dust was smaller than 10 μm in diameter, the extraordinary volume of dust in the air meant that the respirable fraction (particles less than 10 μm) represented a significant amount of exposure. Finally, although only a small percentage of particles larger than 10 μm impact lower airways, the huge magnitude of the WTC dust cloud meant that a small percentage of particles that penetrated deep into the lung may have contributed to a significant exposure. In fact, a study of 39 firefighters from the Fire Department of the City of New York (FDNY) 10 months after exposure[37] demonstrated that the WTC dust did make it down into their lower airways, as particulate matter (>10 μm) consistent with WTC dust, with associated increases in inflammatory cells and cytokines in induced sputum.

UPPER AND LOWER RESPIRATORY DISORDERS

World Trade Center Cough

During the first 5 years post-September 11th, 2001, high rates of upper and lower respiratory irritant symptoms, primarily cough, were described in at least 7 WTC rescue and recovery worker groups (Table 1). 1) In 13,854 previously healthy, exposure-stratified FDNY rescue workers, self-reported daily cough was present in 99% on Day 1 (September 11th, 2001), 53% during the first month postcollapse, 46% during the first year postcollapse, and 31% during the next 2–4 years.[38] 2) in the NY/NJ WTC consortium that follows the non-FDNY cohort of WTC rescue workers and volunteers (police, sanitation, transportation, construction, and others), 69% of the first 9442 responders reported new or worsened upper (62.5% of 9,442) or lower (46.5% of 9,442) respiratory symptoms during their WTC-related efforts, with symptoms persisting to the time of examination in 59% (on average, 8 months after they stopped their rescue, recovery, and cleanup activities).[39] And in another study, this consortium found that in the previously asymptomatic group, 44% developed lower respiratory symptoms during their work at the WTC site. Analysis again demonstrated that the incidence of lower respiratory symptoms was directly related to arrival time at the WTC site.[35] 3) 77% of 240 previously healthy NYC Emergency Service Unit (ESU) police officers had upper and/or lower respiratory symptoms during the first 5 months post-collapse.[40] 4) In 471 NYC police officers (426 with no pre-September 11th chronic respiratory disease), 44% reported having a cough at both 1- and 19-months postcollapse, but over this same time interval, they reported an increasing prevalence of shortness of breath (18.9%–43.6%) and wheeze (13.1%–25.9%).[41] 5) 77% of 96 ironworkers had upper and/or lower respiratory symptoms at 6-month post-collapse.[42] 6) In a study of 269

Table 1. *Upper and Lower Respiratory Disease from World Trade Center Exposures.*

Clinical syndrome	Clinical symptoms	Pulmonary function abnormalities	Study cohort	References
"WTC Cough Syndrome"	Productive cough, wheeze, shortness of breath, chest tightness.	• Bronchial hyperreactivity (methacholine PC20 $\leq$ 8 mg/ml): 23% (high exposure), 8% (moderate exposure). • Decrease in FEV_1, FVC and MFEF 25%–75%, FEV_1/FVC ratio. • Reversible abnormality on spirometry.	FDNY rescue workers; iron workers; ESU police officers; technical, law enforcement, construction workers, transit workers and cleanup workers.	34–50
Chronic Rhinosinusitis or Reactive Upper Airway Dysfunction Syndrome (RUDS)	Nasal congestion/drip, sore throat, sinusitis.	Pathophysiologic testing not available.	FDNY rescue workers; iron workers; ESU police officers; technical, law enforcement, construction workers, transit workers and cleanup workers.	38–40, 42,48,50
Gastroesophageal Reflux Disease (GERD)	Persistent upper gastrointestinal complaints including persistent heartburn.	Pathophysiologic testing not available.	FDNY rescue workers; technical, law enforcement, construction workers, and survivors.	34,39,48, 50,51
Reactive (lower) Airway Dysfunction Syndrome (RADS) or Irritant-induced Asthma	Cough, wheeze, dyspnea.	• Bronchial hyperreactivity (methacholine PC20 $\leq$ 8 mg/ml): 31% (high exposure). • Reversible obstruction.	FDNY rescue workers, rescue and cleanup workers, technical, law enforcement, construction workers, ESU police officers, and residents.	34,39,40, 45–47, 51,52
Pulmonary Function Abnormalities	With or without respiratory symptoms.	• Accelerated FEV_1 and FVC decline. • FEV_1 < LLN and/or FVC < LLN and/or FEV_1/FVC < LLN and/or bronchodilator response. • Large airway dysfunction on forced oscillation. • Persistent bronchial hyperreactivity (methacholine PC20 $\leq$ 8 mg/ml) 1, 3, 6 mo and 2 yr after WTC collapse.	FDNY rescue workers; ESU police officers; iron workers.	34,35,39, 40,42, 50,51
Eosinophilic Pneumonitis	Cough with blackish sputum, fatigue, malaise, fever, dry cough, pleuritic chest pain.	Adult respiratory distress syndrome requiring mechanical ventilation. After corticosteroid treatment and discharge from ICU, hypoxia resolved and pulmonary functions gradually returned to lower limits of normal.	FDNY firefighter (case report).	53
Sarcoid-like Granulomatous Pulmonary Disease (SLGPD) or Granulomatous Pneumonitis	Spectrum of asymptomatic to symptoms of asthma, constitutional symptoms and/or arthralgias.	Biopsy evidence of sterile nonnecrotizing granulomatous disease. Pulmonary functions ranging from normal to documented hyperreactivity. Rare cases of decreased diffusion.	FDNY rescue workers and a case report of an Engineer.	54,55
Bronchiolitis Obliterans	Cough, wheeze, shortness of breath.	Decreasing lung volumes with biopsy evidence for bronchiolitis obliterans and rare nonnecrotizing granuloma.	NYC Police Officer (case report).	56

transit workers, those caught in the dust cloud had significantly higher risk of persistent lower respiratory and mucous membrane symptoms.[43] (7) In 183 cleanup workers, the prevalence of upper and lower respiratory symptoms increased as the cumulative number of days spent at WTC increased.[44] Respiratory consequences have also been noted in WTC studies on community residents, children, and office workers in lower Manhattan.[45–47] A WTC Health Registry study confirmed that of 8418 adults who were caught in the collapse on September 11th, 57% experienced new or worsening respiratory symptoms afterward.[48]

"The World Trade Center Cough Syndrome" was first reported by the FDNY Bureau of Health Services after September 11th, in an article published in the New England Journal of Medicine in 2002[34] and is a chronic cough syndrome, thought to be a consequence of persistent bronchitis (usually asthmatic), rhinosinusitis, gastroesophageal reflux, or any combination of the 3. During the first 6 months following the WTC attack, FDNY described a syndrome of clinical, physiologic, and radiographic abnormalities associated with airway inflammation in an initial cohort of 332 FDNY rescue workers. Because so many were affected, the case definition specified a persistent cough severe enough to require at least 4 weeks of continuous leave (medical, light duty, or retirement) with onset during the 6 months after the WTC collapse. More recently, FDNY reported that between September 11th, 2001 and June 30th, 2007, a total of 1847 (~ 13%) FDNY rescue workers had met this strict case definition and more than 728 have qualified for permanent respiratory disability benefits on the basis of persistent abnormal results on pulmonary function tests including, when indicated, methacholine challenge testing.[38] Clinical symptoms were consistent with aerodigestive mucosal inflammation (rhinosinusitis, bronchitis, acid reflux), with a high rate (87%) of gastroesophageal reflux complaints.[34] Physiologic evidence of asthmatic airway inflammation in those with the syndrome included response to bronchodilators (63% of those diagnosed) and nonspecific bronchial hyper-reactivity determined by methacholine challenge testing (24% of those diagnosed), indicating that these FDNY rescue workers had a high rate of asthmatic physiology.[34] Radiological confirmation of airway inflammation in these firefighters included CT scan evidence of air trapping (abnormal retention of air in the lungs after expiration) in 51% and bronchial-wall thickening in 24%.[34] The incidence of this syndrome was correlated with WTC dust exposure intensity (estimated by initial arrival time at the WTC site). Nearly all of the firefighters and emergency medical technicians (EMTs) who developed WTC cough syndrome had been exposed during the first 48 hours post-collapse, and most had been exposed during the morning of September 11th.[34,38] Recently, consensus treatment guidelines have been published as a joint collaborative effort between the 3 WTC centers of excellence (FDNY, NY/NJ consortium coordinated by Mount Sinai Medical Center and the environmental health Center at Bellevue Hospital), and the WTC Registry.[57]

Upper Respiratory Disease

Reactive Upper Airway Dysfunction Syndrome and Chronic Rhinosinusitis

Reactive upper airway dysfunction syndrome (RUDS) is defined as chronic rhinosinusitis (nasal and/or sinus inflammation) initiated by high-level exposure to inhaled irritants, with recurrence of symptoms after re-exposure to irritants. Diagnosis depends largely on symptoms, and there is no simple way to quantify the severity of the condition. Upper airway symptoms have been described in various occupational groups involved in rescue, recovery, and cleanup at the WTC site (Table 1), with higher prevalence of symptoms in those workers that were more highly exposed. In 13,854 previously healthy FDNY rescue workers, stratified for severity of exposure by arrival time at the WTC site, self-reported chronic sinus congestion and/or drip was reported by less than 5% pre-WTC and was present in 80% on Day 1 (September 11th, 2001), 40% during the first month postcollapse, 25% during the first year postcollapse, and 32% during the next 2–4 years.[38] In this same group, sore or hoarse throat was reported in 63% on Day 1 (September 11th, 2001), 54% during the first month post-collapse, 46% during the first year post-collapse, and 22% during the next 2 to 4 years.[38] An exposure intensity gradient was evident. In the NY/NJ Consortium of non-FDNY rescue workers and volunteers, 66% of those directly exposed to the dust cloud reported upper respiratory symptoms such as congestion, runny nose, headache, sinus pain, sore throat, ear pain or blockage, hoarse voice, etc.[39] They also found that of WTC rescue workers arriving the afternoon of September 11th, 2001, 62% reported experiencing upper respiratory symptoms.[39] In 96 ironworkers, who were on the rubble pile from the afternoon of September 11th, usually on long shifts without respiratory protection, 52% had persistent sinus complaints, with corresponding physical signs such as nasal mucosinusitis and swollen turbinates in at least 30% of the cohort.[42] In 240 NYPD ESU officers between 1 and 5 months after the collapse, 41% had persistent nasal and/or throat symptoms.[40] The main diagnoses associated with these symptoms are chronic sinusitis or chronic rhinitis, as discussed below, but there is considerable overlap with asthmatic and GERD symptoms, and literature pre- and post-WTC clearly shows that without successful sinus treatment, it is difficult to successfully treat those patients who also have asthma.[57,58]

Gastroesophageal Reflux Disease

Before the WTC disaster, persistent upper gastrointestinal complaints under various diagnostic labels have been described in case reports of long-standing Reactive (lower) airway dysfunction syndrome (RADS)[59] and in a case series of persistent gastrointestinal complaints after exposure to chemical irritants.[51] In the general population, Gastroesophageal reflux disease (GERD) has been well described as a causal or exacerbating factor for airway diseases such as sinusitis, asthma, and chronic cough syndrome.[50,57,58] Among FDNY rescue workers, several studies have now described high rates of reflux disease. Of the first 332 FDNY rescue workers diagnosed with WTC cough, 87% reported GERD and were then offered treatment.[34] In a study of 179 exposure-stratified FDNY rescue workers, 45% of those who were found to be hyper-reactive by methacholine challenge testing (1–6 months postcollapse) reported GERD.[60]

Reports of GERD were not limited to FDNY rescue workers (Table 1), as the NY/NJ consortium of non-FDNY worker/volunteers has also reported that in their first year of operation, 54% of their patients had GERD.[39] The WTC Health Registry reported that of a cohort of 8418 adult survivors caught in the collapse, 23.9% reported heartburn or acid reflux not dissimilar from the general population.[48] The authors' personal experience shows that many responders have persistent symptoms that have required prolonged or even chronic use of medications to control acid production. Although no clear mechanism for the development of GERD has been described in this setting, ingestion of airborne particulate WTC material or particulates cleared from the airways, along with stress, dietary triggers, alcohol, weight gain, and medication use (GERD is increased with antibiotics, theophylline, and oral steroids, medications used for WTC-related conditions) are the presumed causes, often acting in combination. Whether GERD is unique to the WTC exposure or represents a previously unrecognized aspect of inhalational injury in general, whether it marks more severe total dust exposure in conjunction with more severe host inflammatory reaction or exacerbation of prior disease, the mechanism by which highly alkaline dust can cause GERD, and whether this gastrointestinal (GI) syndrome will persist or resolve are all unresolved questions. Consensus treatment guidelines clearly show that without successful GERD treatment, there can be only minimally effective treatment for upper and/or lower respiratory conditions such as sinusitis, asthma, and chronic cough.[50,57,58]

Lower Respiratory Disease

Pulmonary function decline or abnormality was significantly related to WTC exposure intensity (based on arrival time) in FDNY and non-FDNY workers (Table 1). This remained true even after accounting for preexistent disease and/or cigarette smoking.[34,39,60–64] For 12,079 FDNY rescue workers in the first year post-WTC, a significantly greater average annual decline in forced expiratory volume in 1 second (FEV_1) of 372 mL was noted in the first year post-September 11th, 2001, when compared with the normal annual decline of 31 mL found in the 5 years of pre-WTC testing–a substantially accelerated decline in pulmonary function.[64] Similar findings were found for the forced vital capacity (FVC), leading to a normal FEV_1/FVC ratio. Preservation of the FEV_1/FVC ratio can be found in airway obstruction when air-trapping is present. In the NY/NJ consortium report on 8384 non-FDNY workers/volunteers, 28% had abnormal pulmonary function test results.[39] They also found that a low FVC, a measure of lung capacity was 5 times more likely among the nonsmoking portion of their cohort than expected in the general US population (which includes smokers and nonsmokers).[39] Overall, WTC dust exposure intensity was related to lower FVC and a higher rate of pulmonary function test abnormalities,[39] demonstrating that WTC exposure had a substantial impact on lung function. Studies in both cohorts (FDNY and non-FDNY) are currently underway to determine the course of pulmonary function over the next 5 years post-WTC, specifically, whether it has improved, stabilized, or declined, and whether there are differences in clinical course within or between cohorts, and which factors may be predictive of favorable or unfavorable outcome.

RADS and Asthma

Occupational RADS is defined by persistent respiratory symptoms and nonspecific airway hyperreactivity in patients with a history of acute exposure to an inhaled agent (gas or aerosol) and no prior history of allergies, smoking, or asthma.[65] Strictly speaking, RADS can only be diagnosed by demonstrating abnormally brisk or intense lower airway obstruction (measured by spirometry) in response to standard provocations (eg, methacholine, histamine, mannitol, cold air, exercise challenge). However, for practical purposes, RADS can be assumed to be present when there are new episodic respiratory symptoms (chest tightness and cough) with spirometric evidence of lower airway obstruction, especially when the obstruction can be reversed by administration of bronchodilating drugs. Asthma or RADS may progress to irreversible lower airway obstructive disease due to chronic inflammation and airway remodeling.

Although RADS was initially reserved for acute exposures to chemical gases and fumes,[65] its use has been extended by some to include acute and even chronic, exposure to respirables particulates. Others prefer to use the term irritant-induced or occupational asthma for such exposures. WTC studies have allowed us to describe the incidence of bronchial hyperreactivity and RADS (or irritant-induced asthma) after a major disaster and to evaluate its persistence longitudinally in a large cohort.

In a sample of FDNY rescue workers whose bronchial hyperreactivity was measured 6 months after September 11, 2001, those who arrived at the WTC site on September 11 were 7.8 times more likely to experience bronchial hyperreactivity than were those firefighters who arrived at the site at a later date and/or had lower exposure levels.[60] In this study of FDNY, RADS emerged in 20% of those who were highly exposed (present during the morning of collapse) and in 8% of moderately exposed rescue workers (present after the morning of September 11 but within the first 48 hours).[60] Consistent with human observational studies, mice acutely exposed to high levels of WTC particulate matter developed pulmonary inflammation and airway hyperreactivity.[66] Findings in FDNY rescue workers over the first 2 years are consistent with the evolving non-WTC scientific literature, indicating that RADS with documented continuing bronchial hyperreactivity can persist in individuals even after exposure had ceased and even with appropriate therapy.[49,62,63]

Currently, for asthma in general and for WTC-exposed subjects specifically, not enough is understood about the mechanism of disease to know if there are important distinctions (mechanism of occurrence, degree of severity, response to treatment, prognosis, etc.) between RADS, irritant-induced asthma and occupational asthma. All we know is that these are lower airway inflammatory diseases that present with provocability (reaction to airborne irritants, cold air, and exercise) and at least partly reversible airway obstruction (Table 1). In the first year of the NY/NJ Consortium Program for non-FDNY workers/volunteers, it was found that 45% reported symptoms consistent with lower airway disorders, including asthma and asthma variants.[35] The WTC Registry has published its findings on self-reported "newly diagnosed asthma (post-September 11th, 2001) by a doctor or other health professional" in WTC rescue and recovery workers.[52] Of the 25,748 WTC workers without a prior history of asthma, newly diagnosed asthma was reported by 926 workers, for a 3-year incidence rate of 3.6%, or 12 times higher than the expected rate of 0.3% in the general adult population.[54] Increased incidence of newly diagnosed asthma was associated with 1) being caught in the dust cloud on September 11, 2001, 2) earlier arrival time relative to the collapse, 3) work on the pile, and 4) cumulative exposure (especially greater than 90 days). When all of the above factors were adjusted in a multivariate analysis, occupation and work tasks were not significant predictors of risk.[52]

For NYC residents, in a telephone survey performed 5 to 9 weeks after September 11, 27% of known adult asthmatics questioned said that they experienced more severe asthma symptoms in the weeks following the attacks.[45] In a post-September 11, 2001 study of Battery Park residents (located adjacent to WTC) who did not report at prior history of asthma, increased respiratory symptoms were reported, and airway hyperreactivity was demonstrated (methacholine challenge testing).[46] A study of Chinese children found increased asthma medication use in a clinic located near WTC, compared with a clinic located in Queens, despite the finding that the same physicians staffed both clinics and both served children of similar Chinese ethnicity.[47] A survey of Medicaid patients during the first 3 months postcollapse (September 11, 2001 to December 31, 2001) showed a significant increase in self-reports of worsening asthma in both lower Manhattan and western Brooklyn, and in the Medicaid Encounter Data System (MEDS), those reporting worsening of their asthma did increase their utilization of asthma healthcare services.[67]

PARENCHYMAL LUNG DISEASES

Reports have shown a higher than expected rate of sarcoidosis or sarcoid-like granulomatous lung disease in FDNY WTC rescue workers.[68] Sarcoidosis is a disorder of the immune system in which groups of white blood cells congregate together to cause lymph-node enlargement and the formation of small inflammatory nodules called granulomas. Most cases have an unknown cause, but environmental causes of sarcoidosis or sarcoidosis-like granulomatous disease are well established, especially after industrial exposure to beryllium.[53] Any organ can be affected, but the most common is the lung and its intrathoracic lymph nodes. In the first 5 years post-WTC (September 11, 2001 to September 10, 2006), pathologic evidence consistent with new-onset sarcoidosis was found in 26 FDNY rescue workers–all with intrathoracic adenopathy (enlarged lymph nodes) and 6 (23%) with additional disease outside the chest.[68] Thirteen were identified during the first year post-WTC (yielding an incidence rate of 86/100,000) and 13 during the next 4 years (yielding an average annual incidence rate of 22/100,000; compared with 15/100,000 for the FDNY personnel during the 15 years pre-WTC and 5–7/100,000 for a male Caucasian population).[56,68] Early arrival time was *not* a predictor of disease, and cumulative exposure time was not reported, but the number of patients with disease was too small to reliably demonstrate an effect. This abnormally high incidence raises the possibility that unknown causative environmental agents were generated or aerosolized during the WTC collapse/combustion.[68] Thus far, studies of WTC patients with sarcoidosis have not identified definitively which environmental agent(s) may be responsible for this disease, and the role of individual immunologic susceptibility to such exposures remains to be studied. Urine beryllium levels were not increased in WTC FDNY rescue workers with or without sarcoidosis, but specialized T-cell studies remain to be performed.[68]

To date, with the exception of sarcoidosis, interstitial lung diseases have not been reported in any case series or population study of WTC workers, but single-case reports of eosinophilic pneumonia,[55] bronchiolitis obliterans,[69] and granulomatous pneumonitis[70] have been described (Table 1). The lay press has reported at least 4 case fatalities as a result of interstitial pulmonary disease in non-FDNY-WTC-exposed subjects.[71] In addition, the FDNY WTC Medical Monitoring and Treatment Program has identified 2 cases of eosinophilic pneumonitis[55] (both resolved on systemic corticosteroids without reoccurrence) and 4 cases of bilateral pulmonary fibrosis (including 1 fatality and 1 who may require lung transplantation, the latter in a patient with the muscle disease polymyositis; personal communication, D. Prezant). These 6 cases showed no pathologic evidence of granulomatous disease or sarcoidosis, and specialized pathologic analyses have not yet been performed to determine if any specific contaminant (silica, heavy metal, etc) was present.

PULMONARY MALIGNANCIES

Because of various carcinogenic compounds (eg, dioxins, PCBs, and PAH) that were found either in the WTC dust or as combustion products from WTC fires,[32,33] there remains the potential for late-emerging malignancies. In the only biomonitoring study, a sample of 321 exposed FDNY firefighters tested 4 weeks after the collapse had measurements of more than 100 analytes, with elevations in only a few (urinary PAH metabolite, antimony, and 2 dioxin congeners) that reached statistically significant differences from background, all falling far short of levels associated with clinical illness.[72] In addition, more than 10,000 FDNY rescue workers were tested for urine beryllium, urine mercury, serum lead, and serum total PCB. Only 1 individual showed significant mercury elevation (thought to be unrelated to WTC), and fewer than 50 had elevated PCB levels, none of which was clinically significant.[38] Samples of sedimented WTC dust from surfaces have shown the presence of asbestos fibers at various amounts and PAHs adherent to particulates.[32] As yet, there has not been any correlation shown between WTC exposure and an increased cancer rate of any kind. Because cancers are latent diseases that can develop long after carcinogen exposure, it is crucial that long-term monitoring be implemented for those who were exposed to the WTC contaminants if we are to ever determine whether cancer rates are altered by WTC exposures.

THE IMPACT OF EXPOSURE TIME ON RESPIRATORY DISEASE

In prior environmental and occupational disasters, much has been made of linking disease to long-term cumulative rather than short-term acute exposures, because diseases such as cancers typically correlate best with cumulative unprotected chronic exposure (eg, mesothelioma and asbestos exposure). However, increased rates of disease have been reported after short-term, high-intensity asbestos exposures.[73] The critical role of long-term exposure does not apply to asthma and sinus conditions, where disease can result from either a single acute exposure in previously healthy nonallergic nonsmokers (RADS) or from recurrent, relatively short-term exposures (occupational asthma, irritant asthma, or sensitization). Similar findings have also been reported for acute and chronic rhinosinusitis (RUDS). And, of course, exacerbations of previously well-controlled asthma and sinusitis are common after exposure to allergens, irritants, and stress. Initially and over the first 6 years, nearly all WTC-related upper and lower airway illnesses in rescue and recovery workers could be attributed to aerodigestive tract inflammation. Studies have documented increased respiratory symptoms, severe persistent cough ("WTC cough"), persistent airway hyperreactivity, RADS or asthma, and declines in pulmonary function among surviving first responders and WTC rescue/recovery workers.[34,35,38–43,60–64,74] In these studies, there was a significant exposure–response gradient, with declines in respiratory health correlating with an earlier arrival time relative to the collapse of the towers.[34,38,39,35,40–43,60–64] Likewise, surviving occupants, caught in the dust cloud on September 11, 2001, had a statistically significant increase in respiratory symptoms.[48]

For upper and lower respiratory illnesses, given the high volume of aerosolized, respirable dust on September 11, 2001 and the lack of appropriate respiratory protection early on, it is not surprising that arrival time provides the best practical measure for a WTC exposure–response index. For nearly all of the FDNY rescue workers, WTC aerodigestive disease has occurred primarily in those who arrived during the first 48 hours after the collapse, with the greatest incidence in those arriving during the morning of the collapse.[34,38,39,60–64] This is not to say that there is no one in the FDNY WTC cohort with WTC aerodigestive disease whose first exposure occurred more than 48 hours postcollapse. And, in fact, a recent WTC registry study on newly diagnosed (post-September 11, 2001) asthma in rescue and recovery workers showed an effect of cumulative exposure (especially greater than 90 hours) even after statistically controlling initial dust cloud exposure and early arrival time.[52] Aerosolized dust was resuspended during the rescue–recovery operations and during cleanup of surrounding interior spaces, and fires continued to burn until mid-December 2001. Although relatively far less common, occurrences of WTC aerodigestive disease in rescue workers/volunteers whose first exposure was more than 48 hours postcollapse could be explained either by high-level exposures generated by activities that disturbed dust in place, during entry of enclosed, poorly ventilated areas, or by accumulation of repeated low-level exposures over time. Of note, in none of these studies has smoking status been found to be a significant confounder.[34,38,39,35,40–43,60–64]

CONCLUSION

A growing body of literature on disaster-related exposures has developed that describes respiratory health consequences to victims, rescue workers, volunteers and nearby residents. Compared with most occupational exposures, disaster-related exposures are far more acute, are often to a wider range of contaminants, and are more difficult to prepare for. Yet, consequences are similar to many occupational and environmental respiratory diseases. For both occupational and disaster-related exposures, the primary emphasis should be on instituting preventive measures through the use of environmental controls and respiratory protection. However, the disaster environment is difficult to control given the unexpected nature and often overwhelming magnitude of the incident, the immediate imperative to rescue survivors, the lack of or difficulty accessing stockpiled respiratory protective equipment, and the widespread use of volunteers. Even after fit-tested respirators have been provided, there are far greater challenges to their effective use in a disaster than in a controlled occupational environment. Valid issues affecting adherence with respirator use include 1) comfort, especially during prolonged use and extremes of temperature; 2) inability to communicate with coworkers in a potentially dangerous environment; 3) training, and 4) on-site supervision. For example, at the WTC site, FDNY has reported that respirators were not available early on and were not used most of the time even when available.[1,61] The WTC registry recently reported that for rescue/recovery workers who arrived on September 11, 2001 and worked in all subsequent time periods, use of masks or respirators did *not* eliminate the risk for newly diagnosed asthma; however, those delays in the initial use of a mask or respirator were associated with an increased incidence of newly diagnosed asthma.[52]

To better prepare for the next disaster, a multi-pronged effort is needed that begins with an understanding that the immediacy of the event will lead to acute exposures, but that this time period should be limited by a dedicated, system-wide approach to provide (as rapidly as possible) and to wear (as frequently as possible) respiratory protection. A thorough understanding of user difficulties in wearing respirators should prompt a redesign of respirators for this environment, and when this is not possible, then work protocols, especially during the recovery phase, should be adjusted to minimize unprotected exposures. Workers and volunteers should not be allowed on site but, instead, should be used off site as support personnel. Exposures can be reduced but can never be prevented, and, therefore, a robust health program for prescreening, monitoring, disease surveillance, and early treatment should be planned in advance and then rapidly instituted beginning with on-site registration of all workers and volunteers.

REFERENCES

1. Centers for Disease Control and Prevention. Use of respiratory protection among responders at the World Trade Center site–New York City, September 2001. *MMWR Morb Mortal Wkly Rep* 2002; 51(6–8):.
2. Thurston GD, Bates DV. Air pollution as an underappreciated cause of asthma symptoms. *JAMA* 2003; 290: 1915–1917.
3. American Thoracic Society. American Thoracic Society Statement: Occupational contribution to the burden of airway disease. *Am J Respir Crit Care Med* 2003; 167: 787–797.
4. Forastiere F. Fine particles and lung cancer. *Occup Environ Med* 2004; 61: 797–798.
5. Cascio WE. Cardiopulmonary health effects of air pollution: is a mechanism emerging? *Am J Respir Crit Care Med* 2005; 172: 1482–1483.
6. Balmes JR. Occupational exposures as a cause of chronic airways disease. In: Rom WN, Markowitz SB, eds. *Environmental and Occupational Medicine.* 4th ed. Philadelphia, PA: Lippincott, Williams and Wilkins, 2007; 517–524.
7. Brooks SM, Tuncale T, McCluskey. Occupational and environmental asthma. In: Rom WN, Markowitz SB, eds. *Environmental and Occupational Medicine.* 4th ed. Philadelphia, PA: Lippincott, Williams and Wilkins, 2007; 418–463.
8. Piirilä P, Nordman H, Korhonen O, Winblad I. A thirteen-year follow-up of respiratory effects of acute exposure to sulfur dioxide. *Scand J Work Environ Health* 1996; 22: 191–196.
9. Hugel DW, Roe R. Nonspecific airway hyperreactivity in nonsmoking bituminous coal miners demonstrated by quantitative methacholine inhalation challenge. *J Lab Clin Med* 1998; 111: 684–691.
10. Petran M, Cocarla A, Baiescu M. Association between bronchial hyperreactivity and exposure to silicon carbide. *Occup Med (Lond)* 2000; 50: 103–106.
11. Ernst P. Dales RE, Nunes F, Becklake MR. Relation of airway responsiveness to duration of work in a dusty environment. *Thorax* 1989; 44: 116–120.
12. Rom WN. Asbestosis, pleural fibrosis and lung cancer. In: Rom WN, Markowitz SB, eds. *Environmental and Occupational Medicine.* 4th ed. Philadelphia, PA: Lippincott. Williams and Wilkins, 2007; 298–316.
13. Attfield MD, Castranova V, Wagner GR. Respiratory disease in coal miners. In: Rom WN, Markowitz SB, eds. *Environmental and Occupational Medicine.* 4th ed. Philadelphia, PA: Lippincott, Williams and Wilkins, 2007; 345–364.
14. Jalloul AS, Banks DE. The health effects of silica exposure. In: Rom WN, Markowitz SB, eds. *Environmental and Occupational Medicine.* 4th ed. Philadelphia, PA: Lippincott, Williams and Wilkins, 2007; 365–387.
15. Lison DF. Hard metal disease. In: Rom WN, Markowitz SB, eds. *Environmental and Occupational Medicine.* 4th ed. Philadelphia, PA: Lippincott, Williams and Wilkins, 2007; 1039–1046.
16. Cormier Y, Lacasse Y. Hypersensitivity pneumonitis. In: Rom WN, Markowitz SB, eds. *Environmental and Occupational Medicine.* 4th ed. Philadelphia, PA: Lippincott, Williams and Wilkins, 2007; 388–401.
17. Hart MB, Sprince NL, Kline JN. Agricultural dust-induced lung disease. In: Rom WN, Markowitz SB, eds. *Environmental and Occupational Medicine.* 4th ed. Philadelphia, PA: Lippincott, Williams and Wilkins, 2007; 570–581.
18. Williams WJ, Wallach ER. Laser microprobe mass spectrometry (LAMMS) analysis of beryllium, sarcoidosis and other granulomatous diseases. *Sarcoidosis* 1989; 6: 111–117.
19. Kreider ME, Christie JD, Thompson B, et al. Relationship of environmental exposures to the clinical phenotype of sarcoidosis. *Chest* 2005; 128: 207–215.
20. Drent M, Bomans PH, Van Suylen RJ, et al. Association of man-made mineral fiber exposure and sarcoid like granulomas. *Respir Med* 2000; 94: 815–820.
21. Nelson D, Robinson C, Dunham MA, Robinson B. Malignant mesothelioma. In: Rom WN, Markowitz SB, eds. *Environmental and Occupational Medicine.* 4th ed. Philadelphia, PA: Lippincott, Williams and Wilkins, 2007; 317–329.
22. Sherman CB, Barnhart S, Miller MF, et al. Firefighting acutely increases airway responsiveness. *Am Rev Respir Dis* 1989; 140: 185–190.
23. Chia KS, Jeyaratnam J, Chan TB, Lim TK. Airway responsiveness of firefighters after smoke exposure. *Br J Ind Med* 1990; 47: 524–527.
24. Kinsella J, Carter R, Reid W, et al. Increased airways reactivity after smoke inhalation. *Lancet* 1991; 337: 595–597.
25. Baxter PJ, Ing R, Falk H, Plikaytis B. Mount St. Helens eruptions: the acute respiratory effects of volcanic ash in a North American community. *Arch Environ Health* 1983; 38: 138–143.
26. Cavalcanti dos Santos Antao V, Pinheiro GA, Parker JE. Lung diseases associated with silicates and other dusts. In: Rom WN, Markowitz SB, eds. *Environmental and Occupational Medicine.* 4th ed. Philadelphia, PA: Lippincott, Williams and Wilkins, 2007; 525–542.
27. Lockey JE, Kapil V, Wiese NK. Man-made vitreous fibers, vermiculite and zeolite. In: Rom WN, Markowitz SB, eds. *Environmental and Occupational Medicine.* 4th ed. Philadelphia, PA: Lippincott, Williams and Wilkins, 2007; 330–344.
28. Cullinan P, Acquilla S, Ramana Dhara V. Respiratory morbidity 10 years after the Union Carbide gas leak at Bhopal: a cross sectional survey. *BMJ* 1997; 314: 338–343.
29. Dharva RV. The union carbide disaster in Bhopal: a review of health effects. *Arch Environ Health* 2002; 57: 391–404.

30. Dikshit RP, Kanhere S. Cancer patterns of lung, oropharynx and oral cavity cancer in relation to gas exposure at Bhopal. *Cancer Causes Control* 1999; 10: 627–636.
31. Shorr AF, Scoville SL, Cersovsky SB, et al. Acute eosinophilic pneumonia among US military personnel deployed in or near Iraq. *JAMA* 2004; 292: 2997–3005.
32. Lioy PJ, Weisel CP, Millette JR, et al. Characterization of the dust/smoke aerosol that settled east of the World Trade Center (WTC) in lower Manhattan after the collapse of the WTC 11 September 2001. *Environ Health Perspect* 2002; 110(():7) 703–714.
33. McGee JK, Chen LC, Cohen MD, et al. Chemical analysis of World Trade Center fine particulate matter for use in toxicologic assessment. *Environ Health Perspect* 2003; 111: 972–980.
34. Prezant DJ, Weiden M, Banauch GI, et al. Cough & bronchial responsiveness in firefighters at the World Trade Center site. *N Engl J Med* 2002; 347: 806–815.
35. Centers for Disease Control and Prevention. Physical health status of world trade center rescue & recovery workers & volunteers–New York City, July 2002–August 2004. *MMWR Morb Mortal Wkly Rep* 2004; 53: 807–812.
36. Toren K, Brisman J, Hagberg S, Karlsson G. Improved nasal clearance among pulp-mill workers after the reduction of lime dust. *Scand J Work Environ Health* 1996; 22: 102–107.
37. Fireman EM, Lerman Y, Ganor E, et al. Induced sputum assessment in New York City firefighters exposed to World Trade Center dust. *Environ Health Perspect* 2004; 112: 1564–1569.
38. Kelly KJ, Niles J, Corrigan M, et al. FDNY WTC Health Effects–a six year assessment. Fire Department of the City of New York. 9/11/2007. Available on-line at: http://www.nyc.gov/html/om/pdf/2007/wtc_health_impacts_on_fdny_rescue_workers_sept-2007.pdf.
39. Herbert R, Moline J, Skloot G, et al. The world trade center disaster and health of workers; five year assessment of a unique medical screening program. *Environ Health Perspect* 2006; 114: 1853–1858.
40. Salzman SH, Moosavy FM, Miskoff JA, et al. Early respiratory abnormalities in emergency services police officers at the World Trade Center site. *J Occup Environ Med* 2004; 46: 113–122.
41. Buyantseva LV, Tulchinsky M, Kapalka GM, et al. Evolution of lower respiratory symptoms in New York police officers after 9/11: a prospective longitudinal study. *J Occup Environ Med* 2007; 49: 310–317.
42. Skloot G, Goldman M, Fischler D, et al. Respiratory symptoms & physiologic assessment of ironworkers at the World Trade Center disaster site. *Chest* 2004; 25: 1248–1255.
43. Tapp LC, Baron S, Bernard B, et al. Physical and mental health symptoms among NYC Transit Workers seven and one half months after the WTC attacks. *Am J Ind Med* 2005; 47: 475–483.
44. Herbstman JB, Frank R, Schwab M, et al. Respiratory effects of inhalation exposure among workers during the clean-up effort at the WTC disaster site. *Environ Res* 2005; 99: 85–92.
45. Centers for Disease Control and Prevention. Self-reported increase in asthma severity after the September 11 attacks on the world trade center–Manhattan, New York, 2001. *MMWR Morb Mortal Wkly Rep* 2002; 51: 781–784.
46. Reibman J, Lin S, Hwang S, et al. The World Trade Center residents' respiratory health study: new onset respiratory symptoms and pulmonary function. *Environ Health Perspect* 2005; 113: 406–411.
47. Szema AM, Khedkar M, Maloney PF, et al. Clinical deterioration in pediatric asthmatic patients after September 11, 2001. *J Allergy Clin Immunol* 2004; 113: 420–426.
48. Brackbill, R. Thorpe L, DiGrande L, et al. Surveillance for world trade center heath effects among survivors of collapsed and damaged buildings. *MMWR Morb Mortal Wkly Rep* 2006; 55: 1–18.
49. Banauch GI, Dhala A, Prezant DJ, et al. Pulmonary disease in rescue workers at the world trade center site. *Curr Opin Pulm Med* 2005; 11: 160–168.
50. Irwin RS. Chronic cough due to gastroesophageal reflux disease: ACCP evidence-based clinical practice guidelines. *Chest* 2006; 129: 805–945.
51. Lieberman AD, Craven MR. Reactive intestinal dysfunction syndrome (RIDS) caused by chemical exposures. *Arch Environ Health* 1998; 53: 354–358.
52. Wheeler K, McKelvey, Thorpe L, et al. Asthma diagnosed after September 11, 2001 among rescue and recovery workers: findings from the World Trade Center registry. *Environ Health Perspect* 2007; 115: 1584–1590.
53. Culver DA, Newman LS, Kavuru MS. Gene-environment interactions in sarcoidosis: challenge and opportunity. *Clin Dermatol* 2007; 25: 267–275.
54. Reed CE. The natural history of asthma. *J Allergy Clin Immunol* 2006; 118: 543–548.
55. Rom WN, Weiden M, Garcia R, et al. Acute eosinophilic pneumonia in a New-York city firefighter exposed to World Trade center dust. *Am J Respir Crit Care Med* 2002; 166: 797–800.
56. Prezant D, Dhala A, Goldstein A, et al. The incidence, prevalence and severity of sarcoidosis in New-York City firefighters. *Chest* 1999; 116: 1183–1193.
57. Friedman S, Cone J, Eros-Sarnyai M, et al. Clinical guidelines for adults exposed to World Trade Center Disaster (Respiratory and Mental Health). City Health Information, NYC Department of Health and Mental Hygiene. September 2006, Available online at: http://www.nyc.gov/html/doh/downloads/pdf/chi/chi25-7.pdf.
58. Pratter MR. Chronic upper airway cough syndrome secondary to rhinosinus diseases (previously referred to as postnasal drip syndrome): ACCP evidence-based clinical practice guidelines. *Chest* 2006; 129: 635–715.
59. Demeter SL, Cordasco EM, Guidotti TL. Permanent respiratory impairment and upper airway symptoms despite clinical improvement in patients with reactive airways dysfunction syndrome. *Sci Total Environ* 2001; 270: 49–55.
60. Banauch GI, Alleyne D, Sanchez R, et al. Persistent bronchial hyperreactivity in New York City firefighters & rescue workers following collapse of World Trade Center. *Am J Respir Crit Care Med* 2003; 168: 54–62.
61. Feldman DM, Baron S, Mueller CA, et al. Initial symptoms, respiratory function & respirator use in New York City firefighters responding to the World Trade Center (WTC) disaster. *Chest* 2004; 125: 1256–1264.
62. Banauch GI, Dhala A, Alleyne D, et al. Bronchial hyperreactivity & other inhalation lung injuries in rescue/recovery workers after the World Trade Center collapse. *Crit Care Med* 2005; 33: S102–S106.

63. Banauch GI, Dhala A. Prezant DJ. Airway dysfunction in rescue workers at the World Trade Center site. *Curr Opin Pulm Med* 2005; 11: 160–168.
64. Banauch GI, Hall C, Weiden M, et al. Pulmonary function loss after World Trade Center exposure in the New York City Fire Department. *Am J Respir Crit Care Med* 2006; 174: 312–319.
65. Brooks SM, Weiss MA, Bernstein IL. Reactive airways dysfunction syndrome. *Chest* 1985; 88: 376–84.
66. Gavett S. Haykal-Coates N, Highfill J, et al. World Trade Center fine particulate matter causes respiratory tract hyperresponsiveness in mice. *Environ Health Perspect* 2003; 111: 981–991.
67. Wagner VL, Radigan MS, Roohan PJ, et al. Asthma in Medicaid managed care enrollees residing in New York City: results from a post-World Trade Center disaster survey. *J Urban Health* 2005; 82: 76–89.
68. Izbicki G, Chavko R, Banauch GI, et al. World trade center "Sarcoid-Like" granulomatous pulmonary disease in New York City fire department rescue workers. *Chest* 2007; 131: 1414–1423.
69. Mann JM, Sha KK, Kline G, et al. World Trade Center dyspnea: bronchioloitis obliterans with functional improvement: case report. *Am J Ind Med* 2005; 48: 225–229.
70. Safirstein BH, Klukowitcz A, Miller R, et al. Granulomatous pneumonitis following exposure to the World Trade center collapse. *Chest* 2003; 123: 301–304.
71. DePalma A. Medical views of 9/11 dust shows big gaps. *NY Times* 2006: 1.
72. Edelman P, Osterloh J, Pirkle J, et al. Biomonitoring of chemical exposure among New York City firefighters responding to the World Trade Center fire and collapse. *Environ Health Perspect* 2003; 111: 1906–1911.
73. Seidman H, Selikoff IJ, Hammond EC. Short-term asbestos work exposure and long-term observation. *Ann N Y Acad Sci* 1979; 330: 61–89.
74. Levin S, Herbert R, Skloot G, et al. Health effect of World Trade Center site workers. *Am J Ind Med* 2002; 42: 545–547.

Mount Sinai Journal of Medicine 75:101–107, 2008

Potential for Diffuse Parenchymal Lung Disease after Exposures at World Trade Center Disaster Site

Jaime Szeinuk, MD,[1] María Padilla, MD,[2] and Rafael E. de la Hoz, MD, MPH[1]

[1]Department of Community and Preventive Medicine and
[2]Department of Medicine, The Mount Sinai School of Medicine, New York, NY, USA

ABSTRACT

Objective: The diffuse parenchymal lung diseases (DPLDs) are a heterogeneous group of disorders that result from damage to the lung parenchyma. While the cause of most DPLDs remains unknown, extensive epidemiological and experimental evidence has linked exposure to environmental toxins to the pathogenesis of some of those diseases. The purpose of this review is to examine the potential relation between exposure to toxins released from the World Trade Center (WTC) collapse on September 11th, 2001 and the development of DPLD based on published evidence up to date.

Methods: We examine such evidence from two points of view, (1) exposure, and (2) histopathogenesis. *Exposure*: Analyses of WTC-dust and particle size demonstrate that some portion of the dust was composed of particles small enough to penetrate deep into the lungs, reaching distal airways and alveoli. The presence of such particles has been confirmed in studies of induced sputum and bronchoalveolar lavage in WTC-exposed firefighters. *Histopathogenesis*: *In vitro* and animal experiments and patient evidence suggest that WTC dust is capable of inducing a pulmonary interstitial inflammatory response.

Results: To date, there have been limited clinical reports documenting the development of diffuse parenchymal responses following exposure to WTC dust. No single common pathologic response has been described. The one common denominator in the reports is that the individuals who developed disease were heavily exposed either during the disaster or during the initial 2-to-3 days following the disaster.

Conclusion: DLPDs are probably associated with heavy or extended exposure to the toxins released at the WTC disaster site. Coupled with the historical experience with exposures to occupational toxins this mandates continued long-term clinical observation of this cohort. ***Mt Sinai J Med 75:101–107, 2008.*** *© 2008 Mount Sinai School of Medicine*

Key Words: World Trade Center disaster, diffuse parenchymal lung disease, interstitial lung disease.

Address Correspondence to:

Rafael E. de la Hoz
WTC Medical Monitoring and
Treatment Program
The Mount Sinai Medical Center
New York, NY, USA
Email: Rafael.delahoz@
mssm.edu

The diffuse parenchymal lung diseases (DPLDs) are a heterogeneous group of disorders that result from damage to the parenchyma of the lung. They manifest a spectrum of histopathology with varying degrees of inflammation and fibrosis. The DPLDs primarily affect the lung interstitium, but air spaces, peripheral airways and vessels, and their respective endothelial linings may also be involved. Extensive epidemiological and experimental evidence has linked exposure to environmental toxins to the pathogenesis of some DPLDs.[1–8] Clinical manifestations of pulmonary disease usually occur several years after the initial exposure (a prolonged latency time).

The DPLDs include several different categories of disease:

1. Disorders of known causes including hypersensitivity pneumonitis and pneumoconioses, collagen disease and/or drug-associated lung disease, including radiation
2. Granulomatous lung disorders
3. Idiopathic interstitial pneumonias including idiopathic pulmonary fibrosis, nonspecific interstitial pneumonia, cryptogenic organizing pneumonia, acute interstitial pneumonia, respiratory

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20025

© 2008 Mount Sinai School of Medicine

bronchiolitis-associated interstitial lung disease, desquamative interstitial pneumonia and lymphocytic interstitial pneumonia

4. Rare forms with distinctive and well-defined clinicopathologic features such as lymphangioleiomyomatosis, pulmonary Langerhan's cell histiocytosis/histiocytosis X and eosinophilic pneumonia[9]
5. Inherited disorders: Hermansky–Pudlak Syndrome, Gaucher's, Fabry's diseases

The strategy to categorize DPLD relies on identification of associated features that can help narrow the options. A careful clinical history must include the following:

1. Assessing the presence of associated disease such as collagen disease, immunodeficiency disorders, malignancy, and treatment for malignant disease
2. A detailed inventory of environmental exposures including smoking, drug use, pets, hobbies, travel, and lifelong occupational exposures, with dates, duration of exposure and an itemized description of work duties and exposures at work
3. A family history of lung disease

On physical examination, crackles or rales and finger clubbing are common. Raynaud's phenomenon, joint swelling or tight skin may suggest collagen disease.[9,10] Serologic testing to exclude collagen vascular disease is essential. Chest radiographs, HRCT scans and pulmonary function tests including spirometry, full lung volumes, and diffusing capacity must be performed. Measurement of oxygen saturation at rest and with effort and the 6-minute-walk test (6MWT) should be obtained at the initial assessment and during follow-up examination.[11] Cardiopulmonary exercise tests could be performed according to clinical needs. Often the diagnostic process may require histologic examination of tissue obtained by bronchoscopy and/or surgical lung biopsies. The correct diagnosis often requires multidisciplinary input or consensus.

Exposure to the airborne toxins that resulted from the attack to and subsequent collapse of the World Trade Center (WTC) has been reported to result in the development of new-onset and persistent airway disease. This has been well documented by surveillance, clinical, and epidemiological data.[12–19]

At present, there are three separate case reports and case series that document development of DPLD[20–22] and two that described the development of distal airway disease with evidence of some involvement of interstitial structures[18,23] following occupational exposures at the WTC disaster site. This review will suggest that pulmonary conditions with prolonged latency, especially DPLD, are likely to occur after exposure at the WTC. We will examine the existing evidence from two main points of view: the histopathogenesis and the exposure.

HISTOPATHOLOGICAL EVIDENCE OF DPLD ASSOCIATED WITH WTC EXPOSURE

Inflammatory markers were studied on induced sputum from nonsmoking New York City (NYC) firefighters exposed to WTC dust some 10 months after the collapse of the WTC towers.[24] The firefighters' exposure duration varied from 1 to 75 days. Induced sputum is a noninvasive alternative to studying inhaled particle matter and the inflammatory response of the lung to inhaled toxins and has yielded similar results to bronchoalveolar lavage (BAL) in silica and hard metal workers and other patients with occupational lung diseases.[25] Induced sputum findings on WTC-exposed firefighters were compared to those in a control population of Israeli hospital workers free from respiratory disease, and also a group of Tel Aviv (TA) firefighters. WTC-exposed firefighters and TA-firefighters had significantly more neutrophils, lymphocytes, and eosinophils and fewer macrophages than hospital workers. The percentage of neutrophils and eosinophils significantly increased with longer exposure at the WTC area, thus suggesting a dose (exposure)-response (inflammation) relationship. WTC-exposed firefighter samples contained a higher percentage of particles of size >2 and 5 μm than TA-firefighters, but no relationship between particle size and length of stay at the WTC area was found. Particles found in WTC-exposed firefighters were irregularly shaped as compared to the smaller, regularly shaped particles found in TA-firefighters. Chemical and mineralogical analyses of these particles revealed metal alloys and oxides consistent with the composition of WTC dust and different from that of non–WTC-exposed firefighters. These mineral particles were seen within the macrophages and epithelial cells of the WTC-exposed firefighters. Higher levels of matrix metalloproteinase-9 (MMP-9) were found in firefighters as compared to nonfirefighters and in WTC-exposed firefighters as compared to the TA-firefighters. MMP-9 plays an important role in neutrophil recruitment to the lungs. It is reliably and reproducibly detectable in induced sputum and has been shown to be elevated in workers exposed to hazardous dust. This elevation was interpreted as evidence for exposure-related immune activation in the lungs, which was persistent 10 months after WTC-dust exposure.

Published case reports of DPLD have been scarce. The first case report concerned a firefighter hospitalized with acute eosinophilic pneumonitis several weeks after WTC exposure.[20] BAL analyses showed significant quantities of fly ash, degraded fibrous glass and asbestos fibers. He had a significant inflammatory response with elevated number of BAL cells recovered (730,000 compared to a normal of <250,000), 70% of which were eosinophils. The eosinophils were not degranulated. Lymphocyte subpopulations (B cells, NK cells, CD8+ cells and CD4+ T cells) were normal, but the alveolar CD4+ lymphocytes exhibited a highly stimulated surface phenotype. IL-5, an eosinophilic chemotactic factor, was elevated. BAL revealed 305 fibers per million alveolar macrophages. Published criteria have established that the finding of more than 1 asbestos-body per million alveolar macrophages has a sensitivity and specificity of 89% for asbestosis.[26] Types of fibers included chrysotile- and amosite-asbestos, chromium, and predominantly silicon-containing fibers that were attributed to degraded fibrous glass. A variety of nonfibrous particles were identified, including fly ash, silica, metal particles, and various silicates. The fact that only uncoated asbestos fibers were found was interpreted as indicative of recent exposure since coated asbestos bodies are expected in more regular, chronic exposure.

The second case report was that of nonsarcoidal granulomatous pneumonitis in a WTC-exposed individual.[21] Lung biopsy of this patient demonstrated noncaseating granulomata with silica, silicate and calcium oxalate particles. The patient was acutely ill and recovered with corticosteroid therapy.

A recent report of surveillance data among NYC firefighters suggested an increased incidence after September 11th, 2001 of "sarcoid-like" granulomatous pulmonary disease, in comparison to previous date. This study included 26 NYC firefighters[22] with "sarcoid-like" granulomatous pulmonary disease. Only three of the cases had a total lung capacity and/or diffusion capacity below 80% of predicted, and extrathoracic disease was present in 23% of the cases. Half of the cases were reported within a year of September 11th, 2001, and the remaining half over the following 4 years. Inclusion into this study required pathology finding of noncaseating granulomas without any evidence of foreign body reaction, malignancy or fungal or mycobacterial infection. The effect of a reporting bias related to WTC-related disability claims was not assessed or described. The incidence of other types of DPLDs was not reported as increased.

Constrictive bronchiolitis has been documented histologically in three WTC-exposed workers[18,23] and suggested radiologically in several others.[16] While this entity is excluded from the histological definition of DPLD, one case reported in a WTC-exposed worker[23] demonstrated chronic bronchiolitis, focal obliterative bronchiolitis and rare nonnecrotizing granulomas in the parenchyma on open lung biopsy. Histological findings in a second case included focal and mild interstitial fibrosis with lymphocyte aggregates but without fibroblastic foci or honeycombing.[18] In a third case, end-expiratory air trapping was documented together with peribronchiolar fibrosis in association with severe restriction, reduced diffusion capacity, and radiologic and histologic interstitial pulmonary fibrotic changes.[18] Thus in these three cases of primarily distal airway disease, there is histopathological evidence suggesting some degree of extension of the inflammatory process more distally into the interstitial space.

TOXICOLOGICAL EFFECTS IN LABORATORY ANIMALS AND CULTURED CELLS

Mice exposed to oropharyngeal aspiration of 100-μg samples of WTC-dust particles with a mass-median aerodynamic diameter (MMAD) of less than 2.5 μm ($PM_{2.5}$) exhibited a slight increase in BAL neutrophils. In addition, this exposure caused significant airflow obstruction in response to methacholine challenge. Lower doses administered directly into the airways (32- and 10-μg) or nasal inhalation of 11 mg/m^3 WTC $PM_{2.5}$ for 5 hours did not induce significant inflammation or airflow obstruction. These results suggest that there may be a threshold dose for respiratory effects of WTC $PM_{2.5}$ in healthy mice. Theoretical calculations indicate that the 100-μg dose in mice corresponds to inhalation of 425 $\mu g/m^3$ WTC $PM_{2.5}$ for 8 hours by humans to achieve a comparable dose in the tracheobronchial region. Such exposures appear to have existed immediately after the collapse of the WTC towers, especially given the lack of sufficient respiratory protection in most rescue personnel. However, most individuals would not be expected to experience such exposures if they were not caught in the dust cloud immediately after the collapse of the towers.[27]

Studies in cultured human alveolar macrophages and type II cells exposed to the $PM_{2.5}$ fraction of indoor and outdoor WTC collected and settled dust showed dose-dependent increases in proinflammatory cytokines such as tumor necrosis factor alpha, interleukin (IL)-8, IL-6 and in gamma-glutamyl transpeptidase (GGT), a membrane-bound enzyme that transfers extracellular precursors of the antioxidant glutathione into the cell cytoplasm where it protects against oxidative tissue damage. The coarse and very coarse fractions of WTC PM caused similar though generally lower responses in these cell types.[28]

EXPOSURE TO WTC DUST

Lioy *et al.*[29] and Offenberg *et al.*[30] have reported on the composition and characteristics of collected settled dust samples in the immediate vicinity of the WTC site. By weight at least 96% of these samples were composed of particles larger than 10 μm in MMAD. Respirable particles including coarse (PM_{10}; particles >2.5 μm and <10 μm MMAD) and fine particles ($PM_{2.5}$; particles <2.5 μm MMAD) were found at much lower proportions (<1% and 0.5%–4% respectively). However, settled dust may not accurately represent the relative proportions of airborne respirable particles following the collapse of the towers. Small particles may coalesce or agglomerate to form larger particles, altering the relative proportions of fine and coarse particles. By one estimation, the ambient fine-particle mass concentration was approximately 500 $\mu g/m^3$ of air on September 12th, 2001,[31] which far exceeds the current US EPA 24-hour air quality standard of 65 $\mu g/m^3$ of air.[32] That estimate was also five times what was found in samples of that duration on a day in 2002 when Canadian forest fires affected the air quality in New York City and New Jersey.[31] Bulk dust samples collected from the vicinity of the WTC area were composed of both fibrous and nonfibrous materials. Fibrous compounds included chrysotile asbestos, glass fibers, and fibers from cotton, wood, and paper (cellulose). Nonfibrous components included gypsum, calcite, bassanite (calcium sulfate hemihydrate), and quartz.[33] Composition of dust collected from indoor locations adjacent to the WTC site was similar.[29,34,35]

Attempts to categorize the exposure of individuals to the attacks and collapse of the towers and to the efforts of rescue, recovery and restoration that continued until the end of December 2001, when the fires were finally extinguished, suggest that there were two significant time-windows for a sufficient exposure that could have caused significant health consequences. The first one corresponds to the initial 5–6 hours after the actual collapse of the towers, when the earliest release and distribution of the pulverized building materials and plume emissions caused by the jet fuel fires occurred. The second one occurred between September 11th and 14th, 2001. During this time, fires were still burning and the highest levels of resuspendable particulate mass could be mobilized from surfaces. This period probably ended with the first post-September 11 rains, which may have decreased the intensity of the fires, washed away the settled resuspendable materials and thus reduced exposure to ambient background levels. However, during the days that followed September 14, intermittent releases of high concentrations of gaseous substances and particulate matter into the local atmosphere within the Ground Zero area still occurred, caused by uncovered fires and the disturbance of settled dust by rescue workers during debris movement and removal. Finally, there was exposure in the indoor environment. Given the facts that conditions of affected residences and buildings were not uniform and that people responded differently in their attitudes to confront such exposures, the estimate of this exposure is virtually impossible in terms of general characterization.[31]

Accounts of the exposure in the above-mentioned published case reports that describe DPLD in WTC-exposed individuals show that all cases had sustained heavy exposure to the collapse, occurring within the two time-window of heavy exposure previously described. The firefighter who developed eosinophilic pneumonia[20] was present at the site on the day of the towers' collapse and continued to work for periods of 16 hours a day for 13 days. He developed acute symptoms as early as 2 days after September 11, but his symptoms worsened sometime around September 22. The patient who developed a granulomatous pneumonitis[21] was also exposed on the day of the collapse of the towers and then returned to the area for 1 day on September 14 and later on October 3, on a more regular basis. This patient worked at a building located one block from the WTC. He initially presented symptoms on October 16. Patients included in the NYC firefighter series of "sarcoid-like" granulomas[22] were all present at the site within the initial 3 days after the collapse of the towers. Ten of the 26 reported patients had arrived in the morning of September 11th, 2001, 14 arrived within the next 36-hours and the last 2 arrived on day 3 after the attack. No specific information as to the duration of the exposure is reported in this article. Fifteen of these 26 patients reported symptoms of asthma and one of chest pain; all the rest were asymptomatic. The NYC highway patrol officer who developed bronchiolitis obliterans[23] was exposed to the WTC dust on the day of the attacks and had continued to work for periods of 16 hours a day for 3 weeks when he first experienced symptoms. The two other cases with peripheral airway disease with extension to the parenchyma had worked at the disaster area for over 80 hours before November 30th, 2001.[18]

DISCUSSION

The development of environmentally induced pulmonary fibrosis can be conceptualized as starting with the inhalation of the causative agent. Physical characteristics of the inhaled particles (e.g. dimensions, solubility), chemical composition, dose, and lung clearance mechanisms (including anatomic and physiologic characteristics of the airways, presence of disease and minute ventilation) contribute to the delivery of the agent to the alveolar spaces and affect the distribution, uptake, and retention of the agent at the lung parenchyma and its potential to cause lung injury.[5] Analyses of WTC dust demonstrated that some percentage of the dust was composed of particles small enough to penetrate deep into the pulmonary system and to reach distal airways and alveoli.[29,30,36] The presence and retention of such particles has been confirmed in studies of induced sputum and BAL in WTC-exposed firefighters.[20,24]

Either by their chemical and/or physical characteristics, inhaled toxic particles activate inflammatory mechanisms at the alveolar level. This may lead to an imbalance in the levels of oxidative stress induced by the inhaled agent and antioxidant protective mechanisms. It may also lead to the generation of cytokines that activate effector cells and culminate in the particles being taken up by alveolar epithelial cells and deposited in the interstitium of the lung. At the interstitial level, the particles trigger the inflammatory cascade, which involves mechanisms of inflammation, coagulation, epithelial, and endothelial repair and others, and finally leads to the release of macrophage-derived growth factors, which induce mesenchymal proliferation and extracellular matrix deposition, both hallmarks of pulmonary fibrosis. Variations in the pattern of immunologic response, probably genetically determined, play a major role on the individual's response to the injury and the final development of persistent damage and disease.[5] As previously reviewed herein, animal and *in vitro* studies and case reports suggest that WTC dust is capable of inducing a pulmonary inflammatory response. Animal models[27] and cell culture models[28] have shown the ability of the dust to induce production of inflammatory mediators. Most importantly, studies of induced sputum and of BAL in firefighters and in patients who developed disease as a result of their exposure confirm the presence of elevated levels of MMPs and ILs and alteration in the number and immunologic properties of pulmonary parenchymal inflammatory cells.[20,24]

In most of the patients in whom a clear environmental cause for DPLD can be identified, clinical history reveals exposure to the toxin for long periods of time before development of symptoms[3,4,6] or for shorter periods of time (as short as one month) if the exposure was unusually heavy.[37] In the case of drug-induced lung disease (except for acute, idiosyncratic reactions), for example, fibrotic lung diseases develop after months-to-years of using the medication and sometimes even years after it has been stopped.[7] Similarly, in cases of hypersensitivity pneumonitis, exposure to the sensitizing agent usually occurred over a long period of time.[8] Almost all of the pneumoconioses present after long latency periods. There is, however, a case report of a truck driver who was heavily exposed to a one-time-only emission of fly ash during an industrial accident who developed acute pulmonary disease with silicotic nodules and parenchymal fibrosis 2 weeks after the exposure.[38] These points illustrate the need to accrue and retain significant amounts of toxins in the lung parenchyma in order to trigger clinical disease. The one common denominator in all cases of post-WTC exposure DPLD published so far is the fact that all affected patients were exposed during the period when environmental concentration of WTC dust was at its highest, i.e. during the day of the attacks and/or up to 3 days after the collapse of the towers. The one unusual feature when comparing WTC-DPLD to other more "traditional" environmentally induced pulmonary interstitial diseases is the relatively short latency period between exposure and symptom development, and the lack of a common histopathological pattern.

It is possible that inhalation of the dust generated during the collapse of the WTC towers could act as an additional stimulus to upregulate fibrotic pulmonary response in individuals with preexisting pulmonary fibrotic conditions. The information about combinations of environmental exposures and predispositions and induction of DPLD is scarce. Evidence of increased pulmonary cell toxicity, especially through generation of reactive oxygen radicals, has been reported in the literature. For example, combined exposure of cultured cells or animals previously treated with herbicides (paraquat) to oxygen, resulted in worse cell and tissue damage than could be induced by the herbicide alone.[39,40] A similar additive deleterious response has been reported upon exposure to nitric oxide and hyperoxia.[41,42] Of greater clinical relevance, oxygen therapy can trigger or accelerate the development of pulmonary fibrosis in patients who have been treated previously with chemotherapy or radiotherapy.[43] Similarly it has been reported that patients with asbestos-related lung disease and cigarette smoking present a worsening degree of radiological abnormalities that may suggest more severe degree of pulmonary fibrosis.[44,45] Therefore, there is at least some theoretical basis to suggest the possibility that WTC dust acts as an aggravating agent to an underlying pulmonary fibrotic process.

Is this sufficient to ascribe causation of DPLD to WTC exposure? So far, most of the reports of illnesses related to WTC exposure have described airway pathology. However, eosinophilic pneumonitis, non-sarcoidal granulomatous pneumonitis, sarcoidosis-like granulomas and bronchiolitis obliterans with elements of extension into the parenchyma have all been reported, and most of these cases were exposed at the WTC site during the first few hours and days after the attack and the collapse of the towers, when the concentrations of inhalable toxicants is presumed to have been highest. A genetically determined but yet unidentified predisposition factor to the development of DPLD may have played a role, and explain why similarly exposed individuals did not develop DPLD. The variability in the reported pathological descriptions remains a challenge at the moment. However, three of the four reported cases involved some type of granulomatous response, which is the expected response of the lung parenchyma when exposed to foreign airborne material.

In conclusion, given the historical experience with environmental and occupational respiratory disease, we know that pulmonary consequences of toxin exposure can present long time after exposure. The reports published so far suggest the development of an inflammatory response to WTC dust with clinicopathological consequences in the pulmonary tissue. Therefore, ongoing surveillance and follow up of the cohort of exposed individuals is warranted. We need to determine whether similar patterns of disease will be observed in other heavily exposed workers or victims of the attack. We need to be vigilant of other disease patterns that could present with longer latency periods in less heavily exposed individuals. Lessons learned from historical experience with exposures to occupational toxicants definitely apply to WTC-exposed population and mandate continued and concerted long-term clinical observation, as well as corroboration by epidemiological data.

ACKNOWLEDGMENT

The authors would like to thank Dr Alvin Teirstein for his valuable help.

REFERENCES

1. Beckett WS. Occupational respiratory diseases. *N Eng J Med* 2000; 342: 406–413.
2. Centers for Disease Control and Prevention. Advanced cases of coal workers' pneumoconiosis–two counties, Virginia, 2006. *MMWR Morb Mortal Wkly Rep* 2006; 55: 909–913.
3. Mossman BT, Churg A. Mechanisms in the pathogenesis of asbestosis and silicosis. *Am J Respir Crit Care Med* 1998; 157: 1666–1680.
4. American Thoracic Society. Diagnosis and initial management of nonmalignant diseases related to asbestos. *Am J Respir Crit Care Med* 2004; 170: 691–715.
5. Taskar VS, Coultas DB. Is idiopathic pulmonary fibrosis an environmental disease? *Proc Am Thorac Soc* 2006; 3: 293–298.
6. American Thoracic Society. Adverse effects of crystalline silica exposure. *Am J Respir Crit Care Med* 1997; 155: 761–765.
7. Camus P, Fanton A, Bonniaud P, et al. Interstitial lung disease induced by drugs and radiation. *Respiration* 2004; 71: 301–326.
8. Churg A, Muller NL, Flint J, Wright JL. Chronic hypersensitivity pneumonitis. *Am J Surg Pathol* 2006; 30: 201–208.
9. American Thoracic Society, European Respiratory Society. American thoracic Society/European respiratory society international multidisciplinary consensus classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med* 2002; 165: 277–304.
10. Springmeyer SC, Furst DE. Initial evaluation of diffuse parenchymal lung disease: is it a connective tissue disease? *Sem Respir Crit Care Med* 1999; 20: 83–90.
11. Flaherty KR, Andrei AC, Murray S, et al. Idiopathic pulmonary fibrosis: prognostic value of changes in physiology and six-minute-walk test. *Am J Respir Crit Care Med* 2006; 174: 803–809.
12. Prezant DJ, Weiden M, Banauch GI, et al. Cough and bronchial responsiveness in firefighters at the World Trade Center Site. *N Engl J Med* 2002; 347: 806–815.
13. Banauch GI, Alleyne D, Sanchez R, et al. Persistent hyperreactivity and reactive airway dysfunction in firefighters at the World Trade Center. *Am J Respir Crit Care Med* 2003; 168: 54–62.
14. Feldman DM, Baron SL, Bernard BP, et al. Symptoms, respiratory use, and pulmonary function changes among New York City firefighters responding to the World Trade Center Disaster. *Chest* 2004; 125: 1256–1264.
15. Herbert R, Moline J, Skloot G, et al. The World Trade Center Disaster and the health of workers: five-year assessment of a unique medial screening program. *Environ Health Perspect* 2006; 114: 1853–1858.
16. Mendelson DS, Roggeveen M, Levin SM, et al. Air trapping detected on end-expiratory high resolution CT in symptomatic World Trade Center rescue and recovery workers. *J Occup Environ Med* 2007; 49: 840–845.
17. Buyantseva LV, Tulchinsky M, Kapalka GMP, et al. Evolution of lower respiratory symptoms in New York police officers after 9/11: a prospective longitudinal study. *J Occup Environ Med* 2007; 49: 310–317.
18. de la Hoz RE, Shohet MR, Chasan R, et al. Occupational toxicant inhalation injury: the World Trade Center experience. *Int Arch Occup Environ Health*, 2008; 81: 479–485.

19. Wheeler K, McKelvey W, Thorpe L, et al. Asthma diagnosed after September 11, 2001 among rescue and recovery workers: findings from the World Trade Center Health Registry. *Environ Health Perspect* 2007; 115: 1584–1590.
20. Rom WN, Weiden M, Garcia R, et al. Acute eosinophilic pneumonia in a New York City firefighter exposed to WTC dust. *Am J Respir Crit Care Med* 2002; 166: 797–800.
21. Safirstein BH, Klukowicz A, Miller R, Teirstein A. Granulomatous pneumonitis following exposure to the World Trade Center collapse. *Chest* 2003; 123: 301–304.
22. Izbicki G, Chavko R, Banauch GI, et al. World Trade Center "sarcoid-like" granulomatous pulmonary disease in New York City Fire Department rescue workers. *Chest* 2007; 131: 1414–1423.
23. Mann JM, Sha KK, Kline G, et al. World Trade Center dyspnea: bronchiolitis obliterans with functional improvement: a case report. *Am J Ind Med* 2005; 48: 225–229.
24. Fireman EM, Lerman Y, Ganor E, et al. Induced sputum assessment in New York City firefighters exposed to World Trade Center dust. *Environ Health Perspect* 2004; 112: 1564–1569.
25. Fireman E, Greif J, Schwartz Y, et al. Assessment of hazardous exposure by BAL and induced sputum. *Chest* 1999; 115: 1720–1728.
26. Roggli VL, Coin PG, MacInteryre NR, Bell DY. Asbestos content of bronchoalveolar lavage fluid: a comparison of light and scanning electron microscopic analysis. *Acta Cytol* 1994; 38: 502–510.
27. Gavett SH, Haykal-Coates N, Highfill JW, et al. World Trade Center fine particulate matter causes respiratory tract hyperresponsiveness in mice. *Environ Health Perspect* 2003; 11: 981–991.
28. Payne JP, Kemp SJ, Dear W, et al. Effects of airborne World Trade Center dust on cytokine release by primary human lung cells in vitro. *J Occup Environ Med* 2004; 46: 420–427.
29. Lioy PJ, Weisel CP. Millerette JR, et al. Characterization of the dust/smoke aerosol that settled east of the World Trade Center in lower Manhattan after the collapse of the WTC 11 September 2001. *Environ Health Perspect* 2002; 110: 703–714.
30. Offenberg JH, Eisenreich SJ, Chen LC, et al. Persistent organic pollutants in the dust that settled across lower Manhattan after September 11, 2001. *Environ Sci Technol* 2003; 37: 502–508.
31. Lioy PJ, Georgopoulos P. The anatomy of the exposures that occurred around the World Trade Center site, 9/11 and beyond. *Ann N Y Acad Sci* 2006; 1076: 54–79.
32. Johnson PRS, Graham JJ. Fine particulate matter National Ambient Air Quality Standards: public health impact on populations in the Northeastern United States. *Environ Health Perspect* 2005; 113: 1140–1147.
33. McGee JK, Chen LC, Cohen MD, et al. Chemical analysis of World Trade Center fine particulate matter for use in toxicological assessment. *Environ Health Perspect* 2003; 111: 972–980.
34. Yiin LM, Millerette JR, Vette A. Comparisons of the dust/smoke particulate that settled inside the surrounding buildings and outside on the streets of southern New York City after the collapse of the World Trade Center, September 11, 2001. *J Air Waste Manag Assoc* 2004; 54: 515–528.
35. Tang KM, Nace CG, Lynes CL, et al. Characterization of background concentrations in Upper Manhattan, New York apartments for select contaminants identified in World Trade Center Dust. *Environ Sci Technol* 2004; 38: 6482–6490.
36. Gavett SH. Physical characteristics and health effects of aerosols from collapsed buildings. *J Aerosol Med* 2006; 19: 84–91.
37. Mossman BT, Ehrlich R, Lilis R, et al. Long-term radiological effects of short-term exposure to amosite asbestos among factory workers. *Br J Ind Med* 1992; 49: 268–275.
38. Shrivastava DK, Kapre SS, Cho K, Cho YJ. Acute lung disease after exposure to fly ash. *Chest* 1994; 106: 309–311.
39. Ody C, Junod AF. Direct toxic effects of paraquat and oxygen on cultured endothelial cells. *Lab Invest* 1985; 52: 77–84.
40. Kehrer JP, Haschek WM, Witschi H. The influence of hyperoxia on the acute toxicity of paraquat and diquat. *Drug Chem Toxicol* 1979; 2: 397–408.
41. Narula P, Xu J, Kazzaz JA, et al. Synergistic cytotoxicity from nitric oxide and hyperoxia in cultured lung cells. *Am J Physiol* 1998; 274: L411–416.
42. Hashida K, Sasaki K, Makino N. Interactions of nitric oxide and oxygen in cytotoxicity: proliferation and antioxidant enzyme activities of endothelial cells in culture. *Free Radic Res* 2000; 33: 147–156.
43. Grahmann PR, Brauer M, Huter L, et al. Respiratory failure and pulmonary fibrosis as a late side effect after chemotherapy-induced by oxygen administration. *Pneumologie* 2005; 59: 763–769.
44. Kilburn KH, Warshaw RH. Severity of pulmonary asbestosis as classified by International Labour Organisation profusion of irregular opacities in 8749 asbestos-exposed American workers. Those who never smoked compared with those who ever smoked. *Arch Intern Med* 1992; 152: 325–327.
45. Kilburn KH, Lilis R, Anderson HA, et al. Interaction of asbestos, age, and cigarette smoking in producing radiographic evidence of diffuse pulmonary fibrosis. *Am J Med* 1986; 80: 377–381.

Standardization of Spirometry in Assessment of Responders Following Man-made Disasters: World Trade Center Worker and Volunteer Medical Screening Program

Paul Enright, MD,[1] Gwen Skloot, MD,[2] and Robin Herbert, MD[3]

[1]The University of Arizona, Tucson, AZ, USA
[2]The Division of Pulmonary and Critical Care Medicine and
[3]Department of Community and Preventive Medicine, Mount Sinai School of Medicine, New York, NY, USA

ABSTRACT

Background: Spirometry is the most commonly used pulmonary function test to screen individuals for suspected lung disease. It is also used for screening workers with exposures to agents associated with pulmonary diseases. Although the American Thoracic Society (ATS) provides guidelines for spirometers and spirometry techniques, many factors are not standardized, so that results from individual pulmonary function laboratories vary substantially. These differences can create substantial difficulties in using data pooled from multiple sites to understand health consequences of disasters that involve exposures to pulmonary toxins. This article describes the approach used to minimize these differences for a consortium of institutions who are providing medical monitoring examinations to World Trade Center (WTC) responders. The protocol improved upon the minimal ATS guidelines.

Methods: Spirometric measurements were obtained before and after use of a bronchodilator. A fourth-generation spirometer was chosen that exceeded ATS spirometer accuracy standards. The accuracy was verified at the beginning of each day of testing. Technologists who performed the spirometry tests were centrally trained and certified and received regular reports on their performance. Reference values and normal ranges were obtained from the National Health and Nutrition Examination Survey (NHANES III) data set. A standardized interpretation flowchart was followed to reduce misclassification rates for airway obstruction and restriction. Patients with spirometric abnormalities were referred for more extensive diagnostic testing.

Results: More than 12,000 spirometry tests were performed during the first examination. The 20 spirometers used at the 6 participating institutions maintained accuracy within 3% for more than 4 years. Overall, more than 80% of the test sessions met ATS quality goals. Spirometry abnormality rates exceeded those obtained for adults who participated in the NHANES III survey.

Conclusions: The program allowed standardization of the performance and interpretation of spirometry results across multiple institutions. This facilitated reliable and rapid diagnosis of lung disease in the large number of WTC responders screened. We recommend this approach for postdisaster pulmonary evaluations in other settings. ***Mt Sinai J Med 75:109–114, 2008.*** *© 2008 Mount Sinai School of Medicine*

Key Words: spirometry, quality control, WTC, occupational screening, disaster.

Address Correspondence to:

Robin Herbert
Department of Community and
Preventive Medicine
Mount Sinai School of Medicine
New York, NY
Email: robin.herbert@mssm.edu

BACKGROUND

Pulmonary specialists and respiratory therapists (RTs) rushed to Ground Zero within weeks following

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20027

© 2008 Mount Sinai School of Medicine

September 11th, 2001, and realized that workers who were inhaling smoke, fumes, and dust would likely suffer symptoms and, potentially, loss of lung function from resulting airway inflammation. They brought portable spirometers to measure lung function. During October 2001, these therapists performed 5-minute spirometry tests for more than 1000 workers in a union hall and gave them the reports with dozens of numbers and a computerized interpretation.[1] About half of the reports indicated a lung function abnormality, consistent with restriction of lung volume or airway obstruction.

In November, copies of the 1000 spirometry reports were scanned and sent for review to a pulmonary function expert at 1 of the top 5 medical institutions in the United States. He viewed all of the flow-volume curves and computer interpretations and was disturbed by the overall poor quality of the tests and the very high misclassification rate. A second opinion from Paul Enright agreed. About half of those who were told that they had restriction (a low vital capacity) actually did not (a false positive). On the other hand, some workers with a chronic cough were told that their spirometry result, were normal (according to the computerized interpretation) but that their flow-volume curves showed airway obstruction (a false negative).

During 2002, the National Institute of Occupational Safety and Health (NIOSH) recommended a spirometry Quality Assurance (QA) program for the World Trade Center (WTC) Worker and Volunteer Medical Screening Program. This program addressed all factors that can cause errors in spirometry tested results, such as instruments, technologists, reference equations, interpretation scheme, and medical decisions based on test results (referrals, diagnoses, and treatment plans). This article describes these factors and the standardization used to minimize misclassification rates. Spirometry results from the first examination have been published.[2]

METHODS

Spirometry Maneuver

Spirometry has been used by doctors for more than 150 years.[3] The maneuver is fairly simple to describe and involves 3 separate phases. The first phase is a complete, deep inhalation, the second is a forceful expiration, and the third is maintenance of exhalation effort for at least 6 seconds. Multiple, precise numbers describing instantaneous flows and relative volumes are generated from this maneuver along with a graphic display of all data from the 3 phases. Unfortunately, despite elaborate presentation, in practice, spirometry is often fraught with error.[4–6]

Instruments

Current models of spirometers measure airflow and convert it to exhaled volume (in liters). This allows them to be relatively portable when compared with the older models of volume spirometers, which had to contain up to 10 L of exhaled air and, thus, were rather large.[7] Unfortunately, the compact size of modern spirometers was often at the expense of long-term accuracy. Condensation of moisture from exhaled breath, as well as phlegm coughed into the flow sensors, clogged many models, and falsely increased the measured flows (forced expiratory volume in 1 second or FEV_1) and volumes (forced vital capacity or FVC).[8] Falsely high FVCs were reported by the automated spirometer used in October 2001 for dozens of workers whose fingers partially occluded the flow sensors during the breathing maneuvers. This was not recognized by those who tested Ground Zero workers at that time.

We carefully chose a modern hand-held spirometer with an ultrasonic flow sensor that does not become inaccurate when moisture or phlegm is exhaled into the mouthpiece (EasyOne Diagnostic model, ndd, Switzerland). The mouthpiece is simply an open tube.[9] The American Thoracic Society (ATS) has guidelines for spirometers and spirometry methods,[10] including daily calibration checks to ensure accuracy within 3%. Therefore, the accuracy with which spirometers measured volumes was checked every day by using a 3-L calibration syringe. The 20 spirometers purchased for this project have maintained better than 3% accuracy for the first 4 years, similar to findings of a large multicenter study in Latin America that also used the same spirometer model.[11]

Technologists

The technologist who administers the spirometry test is the most important determinant of test quality.[12] Patients remain passive during most medical tests, but the 3 phases of the spirometry test require athletic-type breathing maneuvers. The technologist must be an enthusiastic "cheerleader" to coax the patient for maximal effort during all 3 phases. Poor coaching causes inaccurate results.

For spirometry performed in occupational clinics, NIOSH certifies 16-hour spirometry courses that train occupational nurses and spirometry technologists.[13] However, there is no review of the

DOI:10.1002/MSJ

quality of spirometry tests performed by a technologist subsequent to the spirometry course. Technologists may choose to take an examination to become a certified pulmonary function technologist (CPFT).[14] However, less than half of the technologists working in hospital-based pulmonary function laboratories in the United States have this credential, because it is not required by most employers. The CPFT exam is not based on actual performance but only on knowledge of spirometry and arterial blood gas tests. Therefore, in our experience, only a minority of respiratory technologists are able to perform good quality spirometry tests. This requires both training and experience.

Technologists for the current program were trained at Mount Sinai Hospital at the inception of this project, and they followed a written manual of procedures. In addition, the spirometer checked the quality of each maneuver and graded the quality of the spirometry test session in real time.[15] However, the technologist could ignore the messages displayed on the spirometer. The quality grade (A–F) was printed on every report for the physician to consider. Each month, Gwen Skloot, the Director of the WTC Pulmonary Function Quality Assurance Program reviewed the quality of all spirometry tests performed by each technologist and reported their success rate in meeting ATS goals for test session quality (equivalent to a quality grade of A or B), as in the multicenter Lung Health Study.[16] Such feedback has been demonstrated to improve spirometry quality.[17] WTC spirometry technologists achieved high success rates throughout this project.

Reference Equations

For 3 decades, a wide range of spirometry reference equations has been available. Predicted values for FEV_1, FVC, and FEV_1/FVC vary according to the reference equation chosen by the spirometer manufacturer as the default or changed by the user according to preference. Choices include Crapo from Salt Lake City, Knudson from Tucson, Morris from Portland Oregon, Miller from Michigan, and many others.[18] None of these investigations included African Americans, Hispanics, or Asian Americans, so that appropriate reference values for these groups were not available. The ATS did not recommend a single source of reference equations for use in the United States until late in 2005,[10] and few pulmonary function test (PFT) labs have adopted them. As a result of variability in reference equations employed, a patient with borderline abnormal lung function could have an interpretation of normal lung function in 1 laboratory, mild restriction in another, and mild airway obstruction in a third laboratory.

From the beginning of the WTC Worker and Volunteer Screening Program, spirometry reference equations from the third National Health and Nutrition Examination Survey (NHANES III) have been used for all sites, including the FDNY. Separate equations for African Americans, Hispanics, and Caucasians are provided, along with the lower limit of the normal range (LLN) for each key spirometry variable, on the basis of the fifth percentile of the healthy subgroup from NHANES III.[19] Although the NHANES III spirometry reference equations were recommended by the ATS PFT standards committee in 2005, this choice was based on opinions of experts on the committee. No comparison was made of differences in the prevalence rates of spirometry abnormalities when using the popular reference equations. These differences are small when testing typical patients evaluated in a hospital-based PFT lab, because most of these individuals have moderate to severe lung disease and are far from the LLN. However, abnormality rates for mild restriction and mild airway obstruction can differ widely when samples of the general population, or blue-collar working groups such as the WTC responders, are compared by using the major reference equations.[20]

RESULTS AND DISCUSSION

Interpretation Scheme

After comparison with a specific set of reference equations, a list of results displayed as a percentage of the predicted value is produced for key spirometry variables FEV_1, FVC, and FEV_1/FVC. However, reference studies do not suggest how various patterns of percentage of predicted results should be interpreted, and several schemes have been published.[21] The physician interpreting results is free to choose any scheme. Therefore, the same set of percentage of predicted results from the same PFT laboratory could inadvertently be interpreted differently depending on which pulmonary specialist interpreted results on that day.[22]

Because inhaled medication for asthma and chronic obstructive pulmonary disease (COPD) is a highly profitable market, there has been financial incentive during the past decade to push the threshold of spirometry abnormality from normal to abnormal and from a pattern of mild restriction (a slightly reduced vital capacity, which is often due to obesity) to a pattern of mild airway obstruction.[23] The industry-sponsored 2001 GOLD guidelines for COPD,[24] for example, consider an FEV_1/FVC <0.70 to indicate airway obstruction, even if the FEV_1 is more than 100% predicted. If the pattern persists after an inhaled bronchodilator, it is considered due to COPD. However, this guideline ignores studies that show that the FEV_1/FVC declines with normal aging;[25] thus, airway obstruction is being underreported in young workers, while the false-positive rate for "COPD" increases dramatically in people older than age 55 years.[26]

The WTC Screening Program used a standardized interpretation method used by CDC NIOSH, which (considering the high rate of respiratory symptoms of WTC responders) assumes a modest increase in pretest probability of airway obstruction. For mild airway obstruction to be interpreted, the FEV_1/FVC must be below the age- and sex-corrected LLN (the fifth percentile, which is about 90% of the predicted value) and the FEV_1 must also be below its LLN (about 80% of predicted). When the FEV_1/FVC is below the LLN but the FEV1 is not, then an interpretation of borderline airway obstruction is given. The FEF 25%–75% is not used to interpret "small airway disease" when the FEV_1/FVC is normal, because this concept has never been validated.[10]

When a patient has airway obstruction, the presence of a low FVC is often called a "mixed pattern" suggesting a superimposed restriction of lung volumes, but about 90% of the time, the total lung capacity (TLC, as measured in a body plethysmograph) is normal, ruling out a restrictive abnormality.[27] The low FVC in such cases is due to early airway closure at the end of forced expiration with resultant air trapping and hyperinflation. The spirometry interpretation algorithm used by the WTC Screening Program does not make the mistake of interpreting this as a "mixed pattern," but instead designates it as a low FVC.

Medical Decisions

Results of medical tests, even when very accurate and reproducible, do not, on their own, establish a diagnosis. Physicians also consider the patient's symptoms, risk factors (such as smoking, family history, and workplace exposures), medication responses, and consequences of falsely positive versus falsely negative diagnoses[28] (i.e., over-diagnosis versus delayed diagnosis, respectively).

A common flaw in medical decisions when evaluating a patient with a history of asthma-like symptoms but with normal spirometry results (lack of airway obstruction or lack of a bronchodilator response) is to rule out asthma. Mild asthma, by definition, is intermittent, and spirometry will usually be normal between exacerbations.[29] If the diagnosis is needed quickly in such patients, a methacholine challenge test[30] or an exhaled nitric oxide test[31] can be performed.

On the other hand, another common mistake when evaluating spirometry results in adult smokers (with or without respiratory symptoms) is to attribute mild airway obstruction to COPD. The patient should be given an inhaled bronchodilator. Ten minutes later (post-bronchodilator adminstration), if the airway obstruction is no longer present, COPD is ruled out, and asthma should be considered.[32] If the patient still has airway obstruction, a diffusing capacity (DLCO) test should be ordered. A low DLCO (corrected for anemia, if present) in an adult smoker with moderate to severe airway obstruction post-bronchodilator adminstration, is highly likely to be due to COPD.[21] However, other causes for fixed airway obstruction are possible, such as cystic fibrosis in young adults and bronchiolitis obliterans in workers exposed to some types of fumes.[33]

In the absence of airway obstruction, a mildly low FVC (below the LLN) is often called "a restrictive impairment," but the clinical correlates of this pattern are not described in medical literature. About half of patients referred to a PFT lab with this pattern do not have restriction of lung volumes when body plethysmography is performed;[34,35] thus, this pattern is better described as a "nonspecific pattern." If the FVC is normal after bronchodilator (or the slow inspiratory vital capacity is normal), then a restrictive impairment is quickly ruled out. If the FVC is moderately to severely low (below 60% predicted) or the patient has symptoms suggesting an interstitial lung disease, such as dyspnea on exertion or a chronic cough, then a DLCO test and chest x-ray (or perhaps a high-resolution lung CT scan) should be ordered.

Challenges

This program was established at a time when there was concern about potential health consequences due to inhalational exposures to a poorly characterized and unprecedented mix of aerosolized chemicals, combustion products, and micronized building materials. Challenges in assessment of the impact of the disaster response effort on respiratory health at the time of program inception included:

1. Limited information on the nature of exposures,
2. Inability to predict the most likely outcomes (given the limited toxicological information),
3. Absence of baseline (pre-exposure) pulmonary function testing and absence of a control population,
4. Provision to individual responders of meaningful clinical evaluations while also obtaining data

to permit population-based study of respiratory outcomes following an unprecedented disaster,

5. Rapid establishment of a multicenter program, staffed primarily by occupational and internal medicine physicians,
6. Rapid creation of standardized protocols and procedures, with clear guidelines about PFT performance and interpretation for health care providers with variable knowledge about performance and interpretation of PFTs,
7. Emphasis on the importance of using standardized equipment and quality assurance (QA) approaches with which no provider had prior familiarity,
8. Interpretation schema (LLN versus the traditional predicted percentage) that was just being introduced in the pulmonary world and with which most occupational medicine providers and internists were unfamiliar,
9. Logistical, administrative, staffing and other operational challenges faced by each clinical center in their efforts to rapidly expand clinical capacity to accommodate a complex and multifaceted clinical operation.

Thus, it was critical to include pulmonologists who had expertise in standardization and QA in PFT performance/interpretation, as well as a dedication to regularly communicate the importance of standardization and QA to healthcare professionals who were striving to establish and provide programs that were patient friendly, while simultaneously addressing a broad spectrum of possible physical health and mental health outcomes.

Despite many challenges, the WTC Screening Program was successful in incorporating a rigorous PFT QA program into a rapidly developed screening program for disaster responders. To minimize the risk of high misclassification rates and inconsistent interpretations present in other settings, this program standardized all steps in spirometry testing and interpretation and maintained a proactive QA program. This facilitated reliable and rapid diagnosis of lung disease in the 12,000 WTC responders who were screened.

SUMMARY

We have demonstrated the feasibility of incorporating a rigorous program of standardized PFT performance, interpretation, and QA into a setting of postdisaster screening, where, by definition, extensive planning is not possible. This model has two major strengths: (1) it has proven to be logistically feasible for use in the difficult period of postdisaster health response, and (2) it has yielded high-quality PFTs for use in both individual clinical care and epidemiologic study of disaster-related health effects.

ACKNOWLEDGMENTS

This work was supported by the Centers for Disease Control and Prevention and the National Institute for Occupational Safety and Health, contract 200-2002-0038 and grant 5U1O-0H008232. This information has not been previously presented or published.

REFERENCES

1. QRS Spirometer. http://www.qrssys.com/233462.ihtml Accessed September 25, 2007.
2. Herbert R, Moline J, Skloot G, et al. The World Trade Center disaster and the health of workers: five-year assessment of a unique medical screening program. *Environ Health Perspect* 2006; 114: 1853–1858.
3. Hutchinson J. On the capacity of the lungs and on the respiratory functions, with a view to establishing a precise and easy method of detecting disease by the spirometer. *Med Chir Tran* 1846; 29: 137–252.
4. Hankinson JL. Pulmonary function testing in the screening of workers: guidelines for instrumentation, performance, and interpretation. *J Occup Med* 1986; 28: 1081–1092.
5. Kunzli N, Ackermann-Liebrich U, Keller R, et al. Variability of FVC and FEV1 due to technician, team, device, and subject in an 8 center study. *Eur Respir J* 1995; 8: 371–376.
6. Ng'ang'a LW, Ernst P, Jaakkola MS, et al. Spirometric lung function: distribution and determinants of test failure in a young adult population. *Am Rev Respir Dis* 1992; 145: 48–52.
7. Hankinson JL. State of the art of spirometric instrumentation. *Chest* 1990; 97: 258–259.
8. Townsend JC, Hankinson JL, Lindesmith LA, et al. Is my lung function really that good? Flow-type spirometer problems that elevate test results. *Chest* 2004; 125: 1902–1909.
9. ndd Easy One spirometer specifications. http://www.ndd.ch/English/Products/EasyOne_fs.html Accessed August 17, 2007.
10. Miller MR, Hankinson J, Brusasco V, et al. Standardization of spirometry. *Eur Respir J* 2005; 26: 319–338.
11. Perez-Padilla R, Vazquez-Garcia JC, Marquez MN, et al. The long-term stability of portable spirometers used in a multinational study of the prevalence of chronic obstructive pulmonary disease. *Respir Care* 2006; 51: 1167–1671.
12. Enright PL. How to make sure your spirometry tests are of good quality. *Respir Care* 2003; 48: 773–776.
13. NIOSH/CDC. What is the NIOSH-approved spirometry training course? http://www.cdc.gov/niosh/topics/spirometry/ Accessed August 17, 2007.
14. NBRC. The Entry Level CPFT Examination given by the National Board of Respiratory

Care. http://www.nbrc.org/ExamsCPFT.htm Accessed August 17, 2007.

15. Banks DE, Wang ML, McCabe L, et al. Improvement in lung function measurements using a flow spirometer that emphasizes computer assessment of test quality. *J Occup Environ Med* 1996; 38: 279–283.
16. Enright PL, Johnson LR, Connett JE, et al. Spirometry in the Lung Health Study. 1. Methods and quality control. *Am Rev Respir Dis* 1991; 143: 1215–1223.
17. Malmstrom K, Peszek I, Botto A, et al. Quality assurance of asthma clinical trials. *Control Clin Trials* 2002; 23: 143–156.
18. American Thoracic Society. Lung function testing: selection of reference values and interpretive strategies. *Am Rev Respir Dis* 1991; 144: 1202–1218.
19. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999; 159: 179–187.
20. Celli BR, Halbert RJ, Isonaka S, Schau B. Population impact of different definitions of airway obstruction. *Eur Respir J* 2003; 22: 268–273.
21. Pellegrino R, Viegi G, Brusasco V, et al. Interpretative strategies for lung function tests. *Eur Respir J* 2005; 26: 948–968.
22. Raghunath AS, Innes A, Norfolk L, et al. Difficulties in the interpretation of lung function tests in the diagnosis of asthma and chronic obstructive pulmonary disease. *J Asthma* 2006; 43: 657–660.
23. Enright P. Flawed interpretative strategies for lung function tests harm patients. *Eur Respir J* 2006; 27: 1322–1323.
24. Pauwels RA, Buist AS, Calverley PMA. Global strategy for the diagnosis, management, and prevention of COPD. *Am J Respir Crit Care Med* 2001; 163: 1256–1276.
25. Quanjer PH, The Pulmonaria Group. Summary equations for spirometry LLN values, based on quality reviews. www.spirxpert.com/GOLD.html accessed April 11, 2007.
26. Hansen JE, Sun X-G, Wasserman K. Spirometric criteria for airway obstruction. Use percentage of FEV_1/FVC ratio below the fifth percentile, not <70. *Chest* 2007; 131: 349–355.
27. Dykstra BJ, Scanlon PD, Kester MM, et al. Lung volumes in 4774 patients with obstructive lung disease. *Chest* 1999; 115: 68–74.
28. Sox HC. Decision-making: a comparison of referral practice and primary care. *J Fam Pract* 1996; 42(2): 155–160.
29. Ulrik CS, Postma DS, Backer V. Recognition of asthma in adolescents and young adults: which objective measure is best? *J Asthma* 2005; 42(7): 549–554.
30. Crapo RO, Casaburi R, Coates AL, Enright PL. For The American Thoracic Society. Guidelines for methacholine and exercise challenge testing-1999. *Am J Respir Crit Care Med* 2000; 161: 309–329.
31. Taylor DR, Pijnenburg MW, Smith AD, De Jongste JC. Exhaled nitric oxide measurements: clinical application and interpretation. *Thorax* 2006; 61: 817–827.
32. Perez-Padilla R, Hallal PC, Vazquez-Garcia JC, et al. Impact of bronchodilator use on the prevalence of COPD in population-based samples. *COPD* 2007; 4(2): 113–120.
33. Kreiss K, Gomaa A, Kullman G, et al. Clinical bronchiolitis obliterans in workers at a microwave popcorn production plant. *N Engl J Med* 2002; 347: 330–338.
34. Glady CA, Aaron SD, Lunau M, et al. A spirometry-based algorithm to direct lung function in the pulmonary laboratory. *Chest* 2003; 123: 1939–1946.
35. Aaron SD, Dales RE, Cardinal P. How accurate is spirometry at predicting restrictive pulmonary impairment? *Chest* 1999; 115: 869–873.

Mental Health of Workers and Volunteers Responding to Events of 9/11: Review of the Literature

Corey B. Bills, BA,[1] Nancy A. S. Levy, BA,[2] Vansh Sharma, MD,[3] Dennis S. Charney, MD,[3] Robin Herbert, MD,[3] Jacqueline Moline, MD, MSc[3] and Craig L. Katz, MD[3]

[1] Pritzker School of Medicine, University of Chicago and Mailman School of Public Health, Columbia University, New York, NY, USA
[2] New York University and Mailman School of Public Health, Columbia University, New York, NY, USA
[3] The Mount Sinai School of Medicine, Department of Psychiatry and Department of Community & Preventive Medicine New York, NY, USA

ABSTRACT

Background: Disaster workers responding to the events of September 11th were exposed to traumatic events. No study has systematically investigated the diverse mental health status and needs of the heterogeneous population of disaster workers responding to the events of September 11th.

Methods: Using PubMed and Medline and the search terms of "September 11, 2001" or "September 11" or "9/11"or "WTC" or "World Trade Center", the authors reviewed all articles that examined the mental health outcomes of workers at one of the three September 11th crash sites or the Fresh Kills landfill in New York City.

Results: In total, 25 articles met study inclusion criteria, often using different methodologies. The articles described varying degrees of mental health symptomatology, risk factors for adverse mental health outcomes, and utilization of mental health services.

Conclusions: The mental health needs of workers exposed to the events of September 11th ranged from little to no care to pharmacotherapy. A range of risk factors, including exposures at the WTC site and occupational activities, impacted on these needs but the role of specific mental health interventions was less clear. These findings suggest the need for a future program for disaster workers consisting of an accessible mental health treatment service supported by comprehensive postdisaster surveillance and emphasis on pre-disaster mental wellness. A number of areas for further consideration and study were identified, including the need for a more diverse exploration of involved responder populations as well as investigation of potential mental health outcomes beyond post-traumatic stress disorder (PTSD). ***Mt Sinai J Med 75:115–127, 2008.*** © *2008 Mount Sinai School of Medicine*

Key Words: September 11th, disaster, mental health, first responders, disaster workers, review.

Address Correspondence to:

Craig L. Katz
Mount Sinai School of Medicine
Department of Psychiatry
New York, NY, USA
Email: craig.katz@mssm.edu

The September 11th, 2001 attacks on the World Trade Center (WTC) and Pentagon, as well as the plane crash in Shanksville, Pennsylvania, have had a profound effect on the health of Americans.[1] There has also been a good deal of attention paid to those who were directly exposed to the attacks: people who worked in the WTC or the Pentagon, those who live or work in downtown Manhattan, those who lost a loved one in the attacks, and first responders who assisted in the rescue and recovery efforts.

Extensive literature dating back many decades has consistently documented that both manmade and natural disasters are associated with psychopathology of varying frequency, duration, and nature in people exposed to these events.[2–4] Typically, the psychological aftermath of a disaster is conceptualized as arising out of a combination of the exposure itself combined with other factors such as prior psychiatric history and extent of postdisaster psychosocial support. Common psychopathological outcomes include

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20026

© 2008 Mount Sinai School of Medicine

major depression, post-traumatic stress disorder, and possibly substance use problems.[4] Intentional acts of violence, such as those that occurred on September 11th, are often cited as associated with greater psychological distress than natural or unintentional disasters.[5] Disasters of mass violence are difficult for victims to comprehend and assimilate, which creates feelings of helplessness and anxiety, and exposes one's vulnerabilitites.[3] On the other hand, others have debated whether manmade or intentional events do in fact lead to greater psychological distress than natural occurrences or instead to different types of distress and concerns.[2]

As their effects radiate out, disasters like September 11th create a community of affected individuals, what Wright *et al.* refers to as a "disaster community" touched by catastrophe.[6] This community consists of progressively broader rings of affected populations radiating out from a central core of the bereaved next of kin. Disaster workers and volunteers such as those who responded to the September 11th terrorist attacks are a prominent component of this overall community, encompassing what Wright called the "service" and "support" providers whose membership in the disaster community lies just beyond the ring of the bereaved. Disaster workers deserve special attention because they are highly exposed members of the disaster community and are thus at risk for developing mental health sequelae as a result of their efforts.[4,6] They will be the focus of this paper.

A number of studies have examined the impact of disaster work on responders. Disaster and rescue workers who responded to the fatal crash of a United Airlines flight experienced higher rates of acute stress disorder immediately after their involvement 13 months later compared with a nonexposed comparison group.[7] Volunteer disaster workers working with the dead after an explosion on the *USS Iowa* were found to be more likely to have PTSD as well as other symptoms of distress if they identified with the deceased as a friend.[8] Five months following a major building explosion in North Carolina, mild distress responses were noted among first responders; 80% of first responders had at least one symptom of PTSD and 10% had eight symptoms, with on-site staff having more symptoms than hospital-based staff.[9] On the other hand, studies of firefighters who responded to the Oklahoma City bombing and of police officers involved in recovery efforts after an oil rig explosion did not find significant elevations in rates of PTSD or psychological distress.[10,11]

In light of this literature, an examination of the mental health effects of September 11th in particular on disaster workers seems especially important for a number of reasons. First, September 11th responders were subjected to severe, ongoing exposure: some witnessed the scene in the immediate aftermath, some lost coworkers in the disaster, some worked directly with bereaved or searching family members, and many of them stayed at the WTC site for many months, working long shifts in the recovery effort. Second, perhaps as many as 40,000 people have been estimated to have participated in rescue and recovery efforts at Ground Zero in New York City, which reflects a responder population many magnitudes larger than anything previously reported. Finally, whereas prior disaster worker literature has involved isolated reports from disparate disasters; in contrast, the scope and duration of the September 11th rescue and recovery response provides an opportunity to examine the effects of one particular event in one community on many responder populations.

In fact, studies on September 11th have examined diverse populations of workers, such as American Red Cross employees, as well as various types of psychopathology, such as PTSD or alcohol abuse. The purpose of this paper is to review the published literature on the mental health of workers responding to the September 11th disasters. By synthesizing what is known about this population, we hope to shape a more extensive body of information about disaster worker's mental health than has been available to date. In doing so, we hope to inform future research and planning toward mitigating the mental health impact of disasters on responders in the future.

METHODS

Studies relating to the events of September 11th, 2001 were deemed eligible for this study if they included the following: original articles or abstracts published between September 2001 and January 2008, inclusion of study subjects considered workers or volunteers at any one of the three crash sites (New York City, NY, World Trade Center; Arlington, VA, The Pentagon; Shanksville, PA), and an analysis of outcomes and exposures related to mental health. There were no restrictions in terms of language, sample size or study design. Studies that did not include analyses of mental health were not included in the study; nor were studies that included individuals that worked at any one of the three sites prior to September 11th, 2001, but not afterwards. Thus, a worker was defined as any individual who spent time at the three crash sites, or at Fresh Kills Landfill (the major collection and sorting site for debris from the WTC in Staten Island, NY), in a capacity to do work-related activities at any point during or after the events of September 11th, 2001.

Articles included in the present review were identified using MEDLINE and the PubMed database of the National Library of Medicine. The following search terms were used: "September 11, 2001" or "September 11" or "9/11"or "WTC" or "World Trade Center". Two independent researchers (CBB and NASL) reviewed this initial list of titles and abstracts to determine whether an article met the parameters for inclusion. Each of these selected articles was read in full to confirm inclusion eligibility. References cited in these papers that had not been retrieved through database searches, were also evaluated for inclusion. Eventual inclusion of articles was based on consensus between the two researchers and a final review of this consensus by a third researcher (CLK).

The following information was extracted from each of the relevant articles: study design and other study characteristics, such as duration of study; type of population included, demographic variables and work-related duties at the crash sites; sample size of cohort; study test measures used (diagnostic and screening tools); exposures to sites, time worked; other characteristics relevant to one's mental status (social and environmental); mental health outcomes of interest; methods of statistical analysis.

RESULTS

Study Background and Subject Demographics

A review done on all articles published between September 2001 and January 2008 revealed 484 articles in total. Of these, only 25 met the inclusion criteria necessary for this study (Table 1). Studies assessed included 19 cross-sectional, 3 cohort, 2 qualitative and 1 case-series. Utility and disaster workers were the focus of eight studies. Loosely defined rescue and recovery workers were included in four studies. American Red Cross (paid and volunteer) staff were the subject of four studies. Other study subjects included New York City (NYC) Police and Fire Department employees, sanitation drivers, NYC transit employees, iron workers, relief workers, mortuary workers, canine search and rescue workers, and other volunteers.

The 25 studies reported on a combined 122,448 individuals, of which 87,341 were defined as workers. Eight studies also provided comparative assessments of 5,260 unexposed workers.

Of the 16 studies that assessed age, the average age ranged from 35.6 to 56.8 years old. The great majority of study subjects were of white males. In 10 of 17 studies the majority of subjects were male, and in 14 of 15 studies the majority of subjects were white.

Among the 25 studies, 9 provided data on education status. Percentages of subjects with some college or more ranged from 36% among utility workers to 79% among a cohort of American Red Cross paid and volunteer staff.

Most subjects were married or in a committed relationship. Among 10 studies where this variable was assessed, rates ranged from 51%. to 84%.

Studies used varying diagnostic and screening tools to measure similar symptoms, making comparisons between the studies difficult. Additionally, among studies that used the same measures, modifications were often made to the measure itself. Studies also varied in their scoring procedures and interpretation of same measures. However, these differences do not entirely negate comparability.

Exposures and Mediating Factors

The 25 studies examined a variety of exposures related to September 11th (see Table 2). The majority of articles (n = 17) defined exposure as work related to September 11th, including deployment to one of the three sites and exposure to disaster stimuli. Other factors of interest included: prior mental health treatment, mental health symptomatology, and prior traumatic exposure, including sexual abuse as a child.

Two articles studied specific exposures to morbid or disturbing images of September 11th, such as witnessing one or both planes hitting the towers, seeing people jump from buildings, witnessing the building collapse, seeing bodies or body parts amid the debris. One of these studies examined exposure to emotionally difficult tasks, such as working with survivors and families of the deceased, as well as exposures to personal threats such as feelings of physical danger or sustaining an injury. One study examined a treatment modality. Specifically, the authors looked at male disaster workers who were exposed to virtual reality therapy for persistent PTSD, as compared with a similar group who received no such intervention.

Findings from Research Studies

The articles reported a host of psychopathological outcomes, including post-traumatic stress disorder

Table 1. *Overview of Cohort in Studies Meeting Criteria for Inclusion in Review.*

No.	Author (date of publication)	Total	Exposed	Un-exposed	Description	Mean Age (±SD)	White (%)	Male (%)	College (%)	Married/Cohabitating (%)
1	Jayasinge, Giosan, Difede *et al.* (2006)	328	328	–	Utility workers	45 (9)	100	71.3	49.7	82.7
2	Gross, Neria, Tao, *et al.* (2006)	1,355	1,131	224	Union and NYC sanitation workers	–	–	–	–	–
3	Evans, Giosan, Patt, *et al.* (2006)	626	–	–	Disaster/utility relief workers	44.7 (9.4)	66.5	93.1	54.6	75.2
4	Dowling, Moynihan, Genet *et al.* (2006)	28,232	–	–	New York City Police Department workers	–	–	–	–	–
5	Jayasinge, Spielman, Cancellare *et al.* (2005)	174	174	–	Utility workers	43.01 (8.77)	67.3	97.6	53	82.8
6	Tapp, Baron, Bernard *et al.* (2005)	269	88	181	NYC transit workers	45	41	85	12.3	–
7	Johnson, Langlieb, Teret *et al.* (2005)	332	332	–	Clean-up and recovery workers	–	–	–	–	–
8	Smith, Katz, Holmes *et al.* (2004)	1,138	–	–	Rescue and recovery workers	41[a]	58	91	–	–
9	Banauch, McLaughlin, Hirschorn *et al.* (2002)	11,336	–	–	Fire Department of New York City workers	–	–	–	–	–
10	Alvarez & Hunt (2005)	114	82	32	Canine SAR handlers	43	96	43	–	73
11	Covell, Essock, Felton *et al.* (2006)	36,672	13,584	–	Rescue and recovery workers	–	–	–	–	–
12	Difede, Cukor, Patt, *et al.* (2006)	17	9	8	Disaster workers with diagnosed PTSD	"mostly middle age"	–	–	–	–
13	Donahue, Jackson, Shear *et al.* (2006)	229	76[b]	153[b]	Project Liberty service recipients	E: 46 (10.9); U: 45.9 (14.7)	E: 76; U: 57	E: 36; U: 31	–	–

14	Elhai, Jacobs, Kashdan, *et al.* (2005)	3,015	408	2,607	American Red Cross paid and volunteer staff	56.8 (12.5)	92.6	36	78.6	67.3
15	Fullerton, Ursano, Reeves *et al.* (2006)	89	–	–	Disaster workers	35.6 (11.9)	85	77	78	57
16	Gaher, Simons, Jacobs *et al.* (2005)	374	–	–	American Red Cross paid and volunteer staff	55.56 (12.24)	92.59	36.37	–	–
17	Galea, Vlahov, Tracy *et al.* (2004)	2,616	280	2334	Adults in New York City tri-state area	–	53.4	46.2	62.8	50.8
18	Katz, Smith, Silverton *et al.* (2006)	8	–	–	Rescue and recovery workers	–	–	62.5	–	–
19	Leck, Difede, Patt *et al.* (2006)	2,122	–	–	Utility workers	45	66.8	100	35.9	73.9
20	Long, Meyer & Jacobs (2006)	3,,055	2667	202	American Red Cross paid and volunteer staff	56.8 (12.5)	93	36	79	67
21	Simons, Gaher, Jacobs *et al.* (2005)	779	–	–	American Red Cross paid and volunteer staff	56.82 (12.07)	92	36	–	–
22	Zimering, Gulliver, Knight *et al.* (2006)	109	–	–	Relief workers	44.1 (9.26)	88.5	85.3	–	84
23	Levin, Herbert, Skloot *et al.* (2002)	97	–	–	Ironworkers	–	–	–	–	–
24	Peterson, Nicolas, McGraw *et al.* (2002)	>400	–	–	Mortuary workers	–	–	–	–	–
25	Perrin, Digrande, Wheeler *et al.* (2007)	28,692	–	–	Rescue and recovery workers	–	–	–	–	–

[a] median age.

[b] 22 exposed (E) involved in rescue effort; 27 unexposed (U) involved in rescue effort.

Table 2. *Study Design, Mediating Factors and Outcomes Measured in Each of the Studies Under Review.*

No.	Study Design	Study Duration (post-September 11th)	Primary Exposure/Risk Factor	Outcome
1	Cross-sectional	12–21 mo	Stressors	Treatment utilization
2	Cross-sectional	20–24 mo	Work conducted on WTC site	PTSD and other psychological sequelae
3	Cross-sectional	21–25 mo	Anger	Distress and social/occupational functioning
4	Cross-sectional	15–27	September 11th, 2001	Behavioral, emotional, physical & cognitive reactions
5	Cross-sectional	–	Physiological and psychological stressors	Treatment utilization
6	Cross-sectional	7.5 mo	Dust exposure & Ground Zero activities	Physical and mental health symptoms
7	Qualitative	6–21 mo	Clean-up and recovery at WTC disaster site	Open-ended response
8	Cross-sectional	10–15 mo	Rescue and recovery work at WTC	Mental health sequelae
9	Cross-sectional	11 mo	Response to WTC attacks	Injuries and illnesses
10	Cross-sectional	7.75 mo	Deployment to the WTC, Pentagon or PA	Psychological sequelae
11	Cross-sectional	24 mo	Rescue and recovery work	Treatment utilization/
12	Cohort	–	VR enhanced treatment	PTSD (measured in CAPS score)
13	Cohort	18–24 mo	Enahnced counseling	Psychological sequelae
14	Cross-sectional	12 mo	Mental health service use prior to September 11th	Treatment utilization
15	Cross-sectional	2 weeks	Disaster exposure to September 11th	Perceived safety
16	Cross-sectional	12–17 mo	Trait negative affect, coping motives	Alcohol-related problems
17	Cross-sectional	6 mo	Involvement in rescue effort	PTSD
18	Case series	–	Ground Zero rescue & recovery work	Mental health needs
19	Cross-sectional	–	Childhood sexual abuse (CSA)	Psychological sequelae
20	Cross-sectional	12 mo	Disaster stimuli	Psychological distress
21	Cross-sectional	12 mo	PTSD symptomatology	Alcohol use
22	Cross-sectional	6–8 mo	Exposure to disaster site	PTSD
23	Cross-sectional	5 mo	Exposure to WTC site for at least 3 days	Physical and psychological symptomatology
24	Qualitative	–	Model for psychological intervention	–
25	Cohort	24–8 mo	Rescue and recovery work at WTC/occupation	PTSD

and post-traumatic stress symptoms (PTSD/PTSS), major depressive disorder (MDD), panic disorder (PD), alcohol abuse and dependence, and generalized anxiety disorder (GAD; see Table 3). Functional impairment was also measured. However, each of the authors used different diagnostic and screening tools and established unique criteria and cut-off points for defining similar outcomes.

Rates of Mental Health Problems among Those Exposed to September 11th

Several articles compared rates of mental health problems in exposed and unexposed workers. A study of Fire Department of New York City (FDNY) rescue workers, found a 17-fold increase in what was described as "stress-related incidents" among workers in the 11 months following the attacks (1,277 vs 75), as compared to the 11 months prior to September 11th.[12]

For articles that compared PTSD prevalence in exposed and unexposed groups, all articles but one[13] found that first responders exposed to September 11th were more likely to suffer from PTSD than their unexposed counterparts.[1,14,15] Rates of depression in those exposed with PTSD (71.8%) and those exposed without PTSD (51.4%) were also statistically significant ($OR = 16.25$; 95% CI: 11.07,23.85, $p < 0.0001$).[14] Another article, comparing two groups of disaster workers, one with indirect and the other with direct exposures to the WTC site found no statistical difference in measures of depressive symptoms (using the State-Trait Personality Inventory-form Y: STPI-Y) between the two groups.[13] Ross notes a current prevalence rate of depression of 16.1% in exposed workers, compared to 4.4% in unexposed workers.[14] Evans *et al.* found an average Beck Depression Index score of 13.9 among exposed workers (a value above the accepted cut-off score of 12).[16,17] Exposure to the September 11th disaster, as well as exposure to the "dust cloud," were positively associated with depressive symptoms, with odds ratios (ORs) of 2.31 (95% CI: 1.04, 5.15) and 2.48 (95% CI: 1.12, 5.51), respectively.[18] One article found a significant difference between rates of unspecified "alcohol use" among exposed and nonexposed workers (6.8% vs 3.9%) workers.[14] In case-control studies of exposed and unexposed workers, one article found a significantly higher prevalence of anxiety among exposed workers (7.2% vs 1.5%);[14] while another article found no significant difference in anxiety scores on the STPI-Y test between workers directly exposed to the WTC and those with indirect exposure.[13]

Perrin *et al.*, relying on data from the WTC Health Registry of the New York City Department of Health and Mental Hygiene, compared PTSD rates among the different occupations involved in the September 11th response and found that PTSD rates were lowest in police officers (6.2%) and highest in unaffiliated volunteers.[26] Performing tasks not common for one's occupation led to increased rates of PTSD.

Many articles reported cross-sectional prevalence rates of mental health problems in exposed worker cohorts. Levin *et al.* reported a rate of psychological distress among exposed iron workers of 70%.[19] Prevalence of PTSD and PTSS among first responder populations ranged from 8%[18] to 22.5%.[20] Only one article reported a cross-sectional depression prevalence, which was 5.6%.[21]

Rates of excessive/hazardous drinking were found to be 3.2% among exposed NYPD officers[22] and 5% among American Red Cross workers.[23] A single article found that 9.5% of first responders replied "yes" to one of the four CAGE questions.[21] The four "CAGE questions" are a widely used screening tool to identify significant alcohol problems based on details of consumption, and answering affirmatively to one of the four questions indicates that the respondent deserves follow-up while positive answers to two or more questions is highly consistent with a significant drinking problem.[24] Gaher and colleagues found low levels of drinking and drinking-related problems in a group of American Red Cross workers, as mean drinks per week equaled 4.92 ($SD = 4.81$). The authors found an association between negative affect and higher number of alcohol-related problems. This association was mediated by the use of alcohol to cope with the situation.[25]

Cross-sectional surveys of anxiety yielded a current prevalence of 5.9%[21] and 11.7%[22] when measured by clinician assessment and self-report, respectively. Patients with depression ($mean = 8.09$, $SD = 2.73$ vs $mean = 10.12$, $SD = 1.87$, $p = 0.002$) and PTSD ($mean = 7.6$, $SD = 2.62$ vs $mean = 10.17$, $SD = 1.83$, $p = 0.0001$) were more likely to have a lower perceived safety.[20] One article found a current prevalence of panic disorder of 5.8%.[21]

Exposures and Mediating Factors

Pre-September 11th

Several pre-September 11th risk factors were studied, including character traits, past events or trauma, or psychosocial supports. Gaher *et al.* studied the relationship between trait negative affect (an intrapersonal variable that indicates high tendencies

Table 3. *Percent of Study Subjects (prior to September 11th and at Current Time of Study) with Mental Health Symptomatology and Mental Healthcare Use.*

	MH Utilization[a]		PTSD/PTSS[a]		MDD[a]		PD[a]		Alcohol Abuse		GAD[a]		Function	
Author (date of publication)	Prior	Current	Prior	Current	Prior	Current	Prior	Current	Prior	Current	Prior	Current	Prior	Current
Jayasinge, Giosan, Difede *et al.* (2006)[a]	20.4	48	–	–	–	26.2	–	26.2	–	–	–	26.2	–	–
Gross, Neria, Tao, *et al.* (2006)	–	–	–	13.5	–	16.1	–	7.2	–	6.8	–	–	–	–
Evans, Giosan, Patt, *et al.* (2006)	–	–	–	12.5	–	–	–	–	–	–	–	–	–	–
Dowling, Moynihan, Genet *et al.* (2006)	–	–	–	–	–	–	–	–	–	3.2[c]	–	–	–	–
Jayasinge, Spielman, Cancellare *et al.* (2005)[b]	28.7	42.5	–	81.6	30.2	42	30.2	42	–	–	30.2	42	–	–
Tapp, Baron, Bernard *et al.* (2005)	–	34[d]	1.9	8	1.5	12	–	–	–	–	–	–	–	–
Smith, Katz, Holmes *et al.* (2004)	3.2	51[e]	–	19.7	–	5.6	–	5.8	–	9.5	–	5.9	–	13–15
Simons, Gaher, Jacobs *et al.* (2005)	–	–	–	–	–	–	–	–	–	5	–	–	–	–
Alvarez & Hunt (2005)	–	–	–	4	–	–	–	–	–	–	–	–	–	–
Elhai, Jacobs, Kashdan, *et al.* (2005)	13.5	10.7	–	–	–	–	–	–	–	–	–	–	–	–
Difede, Cukor, Patt, *et al.* (2006)	–	–	–	100	–	–	–	–	–	–	–	–	–	–
Galea, Vlahov, Tracy *et al.* (2004)	–	–	–	11.9	–	–	–	–	–	–	–	–	–	–
Zimering, Gulliver, Knight *et al.* (2006)	–	–	–	11	–	–	–	–	–	–	–	–	–	–
Perrin, Digrande, Wheeler, *et al.* (2007)	–	–	–	12.4	–	–	–	–	–	–	–	–	–	–

[a] MH = mental health; PTSD = post-traumatic stress disorder; PTSS = post-traumatic stress symptoms; MDD = major depressive disorder; PD = panic disorder; GAD = generalized anxiety disorder.

[b] Combined rate for PTSD, MDD, PD.

[c] Combined rate for excessive alcohol, illegal and prescription drug use.

toward worry, anxiety, and depression) and alcohol coping motives.[25] The authors reported that among American Red Cross workers, negative affect was associated with higher degrees of alcohol problems, especially among younger participants. Leck *et al.* found that among male disaster workers, those exposed to childhood sexual abuse (CSA) were more likely to endorse higher scores on the BDI and the Clinician Administered PTSD Scale.[27]

For transit workers who reported experiencing or witnessing a prior (pre-September 11th) traumatic event, depressive symptoms were significantly associated with both witnessing September 11th (OR = 3.94; CI 95%: 1.27, 12.17) and knowing a WTC victim (OR = 3.75; CI 95%: 1.18, 11.94). Additionally, those with low perceived social support showed a significant association between knowing a WTC victim and post-traumatic stress symptoms (OR = 16.41; CI 95%: 2.14, 722.96).[18] Among canine search and rescue (SAR) workers, those with a previous mental disorder were more likely to report symptoms of depression PTSD, anxiety, or general psychological distress, (Fs (1,60) >10, ps < .05, ds > .67) and they were more likely to meet the criteria for a current diagnosis ($\chi 2$ (1, n = 60) = 5.64, $p < .05$).[15] Dog handlers that were FEMA (Federal Emergency Management Agency) certified, were less likely to report PTSD symptomatology, than those who were not FEMA certified ($p < 0.05$)

September 11th or Peri-event

The vast majority of articles studied exposures to September 11th-related events. Gross and colleagues reported that disaster workers suffered from significantly higher rates of PTSD, depression, panic disorder, and alcohol use when compared with unexposed workers.[14] Among workers exposed to September 11th, those who experienced personal loss or exposure to death or remains were significantly more likely to suffer from PTSD. Among canine SAR workers, those who were exposed to the September 11th tragedy reported more symptoms of psychological distress, PTSD, and peritraumatic dissociation than those who were not exposed to 9/11. Among exposed canine SAR workers, PTSS was significantly associated with high peritraumatic dissociation and low interpersonal support. Anxiety was associated with low interpersonal support and low relationship satisfaction. Psychological distress as measured by the Brief Symptom Inventory (BSI) was significantly associated with high peritraumatic dissociation, low interpersonal support, and low relationship satisfaction.[15]

Among transit workers, exposure to the "dust cloud" was found to be a risk factor for both depressive symptoms and PTSS, although this latter relationship was not significant among women.[18] When perceived social support was controlled for, depressive symptoms were significantly associated with witnessing 9/11 trauma.

Simons *et al.* reported that American Red Cross workers exposed to September 11th who scored higher on the IES-R (indicating greater post-traumatic symptoms) were more likely to report a change in alcohol consumption, greater weekly consumption, and greater hazardous consumption.[23] These associations were stronger for those disaster workers who displayed hyperarousal symptoms (as opposed to avoidance or intrusion symptoms).

A study of disaster relief workers reported that workers developed PTSD symptomatology as a result of direct trauma (i.e. witnessing the attacks) more often than indirect trauma (i.e. hearing survivor narratives). However, there was evidence of indirect trauma causing the symptomatology of PTSD at a rate of 4.6%.[28]

In the WTC Health Registry cohort, earlier start date at Ground Zero and longer duration of work were significant risk factors for development of PTSD in all occupations except police.[26]

Post-September 11th

The impact of post-September 11th treatment modalities was studied by several authors. Virtual reality was found to be effective for the treatment of refractory PTSD in male disaster workers.[29] Among recipients of Project Liberty services, those who received enhanced treatment services (in addition to baseline crisis counseling) reported significant improvement in function, fewer symptoms of depression and grief, and marginally less traumatic stress. This result is relevant to our review as significantly greater proportions of enhanced services recipients were involved in rescue efforts when compared to those who received only crisis counseling.[30] Mortuary workers at the Pentagon exposed to critical incident stress management (CISM) responded positively to the intervention.[31] Workers implemented some of the recommendations suggested by the CISM teams, and there was a very low rate of removal of personnel from mortuary duty. However, this data was entirely qualitative; symptomatology was not measured psychometrically.

Mental Health Services

Of 25 studies, 6 addressed the utilization of mental health services. Among 174 disaster workers deployed to the WTC site, who accepted referrals for psychotherapy, less than half (42.5%) attended at least one psychotherapy session.[32]

In a separate study, Jayasinghe and colleagues report that 48% of 328 utility workers deemed in need of mental health services, accepted mental health psychotherapy referrals after work at the World Trade site. Similar percentages of study participants chose to "consider" (28%) and decline (24%). Decision to accept vs decline mental health screening services was positively associated with current PTSD (OR = 1.12; CI 95%: 1.04, 1.21; $p < .01$), depressive symptoms (OR = 1.07; CI 95%: 1.00, 1.12; $p < .05$) and previous mental health treatment (OR = 2.85; CI 95%: 1.16, 7.03; $p < .05$).[33]

In a cohort of transit workers, 16% agreed when asked if they thought supportive counseling would be helpful.[18] Among rescue and recovery workers accessing medical services at the Mount Sinai School of Medicine, only 2% has previously received mental health services despite 51% having met screening thresholds for significant psychological distress or disorder.[21] In a study of Project Liberty clients, 13% of rescue and recovery workers sought referral for psychotherapy.[34]

Mental health service use among American Red Cross Workers post-September 11th was also assessed.[35] Among all study subjects (n = 3015), 13.5% utilized mental health services prior to September 11th, while only 10.7% accessed similar services after September 11th. Subjects who did not seek mental health services pre-September 11th were 50% more likely to seek mental health services post 9/11, while those who did access services prior to September 11th were 80% less likely. Thus, prior mental health services did not predict individual use post-September 11th.

DISCUSSION

The mental health outcomes of individuals who worked at sites related to the September 11th, 2001, attacks were varied across studies. However, most studies noted substantial rates of PTSD and MDD symptoms. Among the majority of studies, exposure to events at the WTC was associated with greater rates of overall mental health symptomatology. Little research has documented mental health service utilization among WTC workers. No studies to date have assessed the quality of such services and outcomes over time.

Rates of PTSD among individuals who worked at the WTC ranged from 8% in transit workers[18] and 22.5% among a cohort of disaster workers.[20] The difference in rates of PTSD between the two groups is difficult to address, though may correspond to differing levels of exposure. Hypothetically, disaster workers, who worked onsite would have greater exposure over a potentially longer time period to trauma than transit workers whose primary duties were peripheral to the WTC site. However, it is difficult to quantify levels of exposure from each of these two studies.

The range of PTSD risk among September 11th workers were all well above reported lifetime prevalence of PTSD in the general population. In Scandinavia, lifetime prevalence of PTSD is 5.6, while in the United States the risk is slightly higher -7.8.[36,37] Overall, women are far more likely to develop PTSD, as lifetime prevalence among men and women in the United States is 5% and 10.5%, respectively.[37] The majority of rates of PTSD among September 11th workers concerned males, thus the difference among the general population and September 11th workers may be even more pronounced.

However, the range of PTSD did not surpass the estimated risk reported just after the events of September 11th. Based on average risks associated with the Oklahoma City rescue workers, the estimated risk of PTSD among exposed September 11th rescue workers was 24%.[38]

Exposure to the events of September 11th among workers at each of the sites was also associated with MDD. However, this association was mixed, as only one of two studies noted an association between MDD and exposure status.[14] Furthermore, the 5.6% prevalence of MDD among individuals seeking treatment at the Mount Sinai School of Medicine, was little different than the 12-month prevalence rate of 5.28 (among the general US population), reported by Hasin and colleagues.[39]

In the immediate aftermath of September 11th several mental health programs were established with the intention of providing free screening, referral and treatment services. Six studies report on this process. In those studies examining service utilization, September 11th workers generally were reluctant to utilize screening and referral mental health services. Surprisingly, studies assessing treatment use prior to September 11th conflicted on whether previous mental health treatment was predictive of post September 11th treatment use. Many individuals responding to September 11th may have been reluctant to seek services given the larger cultural and social context in which they work.

This review of the literature on the mental health of September 11th first responders and volunteer workers has revealed several avenues for research and planning. Since 2001, 25 studies have addressed the mental health needs of workers at theWTC, encompassing a far larger number of studies and of subjects than has ever previously been available for any other disaster. However, most of this literature was limited to one first responder group (e.g. American Red Cross workers, transit workers, ironworkers, etc) or a single psychiatric outcome (e.g. PTSD). Future research in this area should study a broader cross-section of first responders from various organizations, occupations, and disaster sites and take regard of various psychiatric outcomes. Quantitative outcomes should include not just PTSD but also major depression, alcohol use disorders, general distress, psychiatric symptoms, and psychosocial impact of the disaster and these disorders.

A mental health registry akin to the WTC Health Registry that begins sooner than two years after an event, as did this registry, and follows all responder groups longitudinally and across a span of mental health outcomes would provide the most comprehensive model for future monitoring of mental health status among responders. It would also assure uniformity of data collection methods, while capturing a substantial pool of subjects like that reflected in the 25 studies in this paper.

This disaster mental health registry should also possess ready availability of epidemiological expertise for prompt and expert analysis of data, thereby assuring rapid turnaround of these analyses for shaping available services in "real time." The public health impact of such analyses would be substantial and immediate if the putative surveillance program were paired with an accessible and specialized treatment resource such as that provided at the Mount Sinai School of Medicine.[21,40] In this way, surveillance and research can inform not only future disasters but the current one as well.

While several studies of the 25 reviewed documented factors associated with referral for psychiatric treatment, only a few provide an indication as to whether individuals who are offered referral services partake in treatment. There is no indication as to whether individuals who are actually offered referral services partake in treatment. Further, no studies have assessed the successes of treatment programs with September 11th worker populations over time. Given the paucity of longitudinal research on mental health outcomes among September 11th workers it is difficult to formulate conclusions about best practices for providing care and treatment to these individuals.

To the large scale tandem of surveillance and treatment of disaster workers which we are proposing for future disasters should also be added a third component, a more mentally healthy and robust disaster workforce. The range of exposure and other risk factors associated with workers' mental health responses to September 11th suggests that further study and policy making should be devoted to fostering circumstances favorable to avoiding adverse psychological sequelae from disaster work. These may include considering responders' personal histories and psychological coping styles, devising means of limiting disaster exposure (such as limits on work hours), and promoting protective post-disaster conditions such as interpersonal support. Once again, potentially significant aspects of the post-disaster conditions that warrant special attention are the utilization, nature and outcome of short and long term mental health interventions.

None-the-less, so long as large scale disasters occur at a pace that outstrips research endeavors, the September 11th literature points to several possible measures available for preventing the likelihood of significant mental health sequelae and maximizing worker resilience in the future. The literature from September 11th underscores the mental health benefits of deploying and utilizing responders who are maximally trained and organized and called upon to do what they are trained to do. Ensuring better access to routine mental health services pre-disaster may also be protective. The deployment and utilization of future disaster workers should be guided, as much as possible, by regular monitoring of workers' exposure, including duration of work, in a manner akin to monitoring radiological exposure.

Finally, all but two[40,41] of the 25 articles conducted a quantitative assessment that used some form of psychometric diagnostic or screening tool. While quantitative assessments are useful in assessing the magnitude of psychopathological outcomes among disaster workers, they do little to capture either the experience of the attacks themselves, or the psychological distress that followed. For example, while it is important to study the rate of PTSD in a disaster worker population, it is equally important to evaluate the experience and manifestations of PTSD among disaster workers, i.e. the images that dominate their flashbacks, the noises that trigger a startle response, etc. In addition, standardized psychometric tools such as the BAI or the PCL fail to address the underlying thought processes behind mental illness. Using the BAI to assess a first responder's anxiety level, for example, does not reveal, whether that anxiety is caused by memories of the attack, fear of a future attack, or worries over September 11th-related health problems. As much of the psychotherapeutic enterprise around PTSD in particular relies on a theoretical model wherein avoided and often dissociated traumatic memories need to be re-integrated into a meaningful and less painful state of awareness and recall,[42] so too should the research enterprise better examine the thought processes and experiences underlying symptomalogy.

Qualitative study of psychopathology among this unique group may deepen our understanding of psychological responses to disasters. Qualitative studies, focused on the experience of September 11th as much as the psychological outcomes, will reveal the commonalities of the disaster experience and will allow the mental health community to better understand and serve clients who are recovering from such trauma.

The major limitation of this study is its reliance on 25 studies using often disparate methods of assessment and different sub-populations of September 11th disaster workers. However, its strength lies in its composite view of the mosaic of findings among the largest number of studies about disaster workers from one disaster available in the scientific literature. This composite has yielded a prescription of how best to identify, track, and treat disaster related mental illness in future disaster workers.

ACKNOWLEDGMENT OF GRANT SUPPORT

National Institutes of Occupational Safety and Health (grants U10 OH008232-04 AND U10 OH008225-01) the Robin Hood Foundation Relief Fund, the American Red Cross Liberty Fund, The September 11th Recovery Program, The Bear Stearns Charitable Foundation, and The September 11th Fund.

REFERENCES

1. Galea S, Vlahov D, Tracy M, et al. Hispanic ethnicity and post-traumatic stress disorder after a disaster: evidence from a general population survey after September 11, 2001. *Ann Epidemiol* 2004; 14(8): 520–531.
2. Garakani A, Hirschowitz J, Katz CL. General disaster psychiatry. *Psychiatr Clin North Am* 2004; 27(3): 391–406.
3. Norris FH, Friedman MJ, Watson PJ. 60,000 disaster victims speak: Part II. Summary and implications of the disaster mental health research. *Psychiatry* 2002; 65(3): 240–260.
4. Katz CL, Pellegrino L, Pandya A, et al. Research on psychiatric outcomes subsequent to disasters: a review of the literature. *Psychiatry Res* 2002; 110(3): 201–17.
5. Norris FH, Friedman MJ, Watson PJ, et al. 60,000 disaster victims speak: Part I. An empirical review of the empirical literature, 1981–2001. *Psychiatry* 2002; 65(3): 207–239.
6. Wright KM, Ursano RJ, Bartone PT, Ingraham LH. The shared experience of catastrophe: an expanded classification of the disaster community. *Am J Orthopsychiatry* 1990; 60(1): 35–42.
7. Fullerton CS, Ursano RJ, Wang L. Acute stress disorder, posttraumatic stress disorder, and depression in disaster or rescue workers. *Am J Psychiatry* 2004; 161: 1370–1376.
8. Ursano RJ, Fullerton CS, Vance K, Kao TC. Posttraumatic stress disorder and identification in disaster workers. *Am J Psychiatry* 1999; 156: 353–359.
9. Durham TW, McCammon SL, Allison EJ. The psychological impact of disaster on rescue personnel. *Ann Emerg Med* 1985; 14: 664–668.
10. North CS, Tivis L, McMillen JC, et al. Coping, functioning, and adjustment of rescue workers after the Oklahoma City bombing. *J Trauma Stress* 2002; 15: 171–175.
11. Alexander DA, Wells A. Reactions of police officers to body-handling after a major disaster: a before and after comparison. *Br J Psychiatry* 1991; 159: 547–555.
12. Banauch G, McLaughlin M, Hirschorn R, et al. Injuries and illnesses among New York City Fire Department rescue workers after responding to the World Trade Center attacks. *JAMA* 2002; 288(13): 1581–1584.
13. Long ME, Meyer DL, Jacobs GA. Psychological distress among American Red Cross disaster workers responding to the terrorist attacks of September 11, 2001. *Psychiatry Res* 2006; 149: 303–308.
14. Gross R, Neria Y, Tao XG, et al. Posttraumatic stress disorder and other psychological sequelae among world trade center clean up and recovery workers. *Ann N Y Acad Sci* 2006; 1071: 495–499.
15. Alvarez J & Hunt M. Risk and resilience in canine search and rescue handlers after 9/11. *J Trauma Stress* 2005; 18(5): 497–505.
16. Evans S, Giosan C, Patt I, et al. Anger and its association to distress and social/occupational functioning in symptomatic disaster relief workers responding to the September 11, 2001, World Trade Center disaster. *J Trauma Stress* 2006; 19(1): 147–52.
17. Snijders AH, Robertson MM, Orth M. Beck Depression Inventory is a useful screening tool for major depressive disorder in Gilles de la Tourette syndrome. *J Neurol Neurosurg Psychiatry* 2006; 77: 787–789.
18. Tapp LC, Baron S, Bernard B, et al. Physical and mental health symptoms among NYC transit workers seven and one-half months after the WTC attacks. *Am J Ind Med* 2005; 47(6): 475–483.
19. Levin S, Herbert R, Skloot G, et al. Health effects of World Trade Center site workers. *Am J Ind Med* 2002; 42(6): 545–547.
20. Fullerton CS, Ursano RJ, Reeves J, et al. Perceived safety in disaster workers following 9/11. *J Nerv Ment Dis* 2006; 194(1): 61–63.
21. Smith RP, Katz CL, Holmes A, et al. Mental health status of World Trade Center rescue and recovery workers

and volunteers - New York City, July 2002-August 2004. *MMWR Morb Mortal Wkly Rep* 2004; 53(35): 812–815.
22. Dowling FG, Moynihan G, Genet B, Lewis J. A peer-based assistance program for officers with the New York City Police Department: report of the effects of Sept. 11, 2001. *Am J Psychiatry* 2006; 163(1): 151–153.
23. Simons JS, Gaher RM, Jacobs GA, et al. Associations between alcohol use and PTSD symptoms among American Red Cross disaster relief workers responding to the 9/11/2001 attacks. *Am J Drug Alcohol Abuse* 2005; 31(2): 285–304.
24. Mayfield D, McLeod G, Hall P. The CAGE questionnaire: validation of a new alcoholism screening instrument. *Am J Psychiatry* 1973; 131: 1121–1123.
25. Gaher RM, Simons JS, Jacobs GA, et al. Coping motives and trait negative affect: testing mediation and moderation models of alcohol problems among American Red Cross disaster workers who responded to the September 11,2001 terrorist attacks. *Addict Behav* 2006; 31(8): 1319–1330.
26. Perrin MA, DiGrande L, Wheeler K, et al. Differences in PTSD prevalence and associated risk factors among World Trade Center rescue and recovery workers. *Am J Psychiatry* 2007; 9: 1385–1394.
27. Leck P, Difede J, Patt I, et al. Incidence of male childhood sexual abuse and psychological sequelae in disaster workers exposed to a terrorist attack. *Int J Emerg Ment Health* 2006; 8(4): 267–274.
28. Zimering R, Gulliver SB, Knight J, et al. Posttraumatic stress disorder in disaster relief workers following direct and indirect trauma exposure to Ground Zero. *J Trauma Stress* 2006; 19(4): 553–557.
29. Difede J, Cukor J, Patt I, et al. The application of virtual reality to the treatment of PTSD following the WTC attack. *Ann N Y Acad Sci* 2006; 1071: 500–501.
30. Donahue SA, Jackson CT, Shear KM, et al. Outcomes of enhanced counseling services provided to adults through Project Liberty. *Psychiatr Serv* 2006; 57(9): 1298–1303.
31. Peterson AL, Nicolas MG, McGraw K, et al. Psychological intervention with mortuary workers after the September 11 attack: the Dover Behavioral Health Consultant model. *Mil Med* 2002; 167(suppl. 9): 83–86.
32. Jayasinghe N, Spielman L, Cancellare D, et al. Predictors of treatment utilization in world trade center attack disaster workers: role of race/ethnicity and symptom severity. *Int J Emerg Ment Health* 2005; 7(2): 91–99.
33. Jayasinghe N, Giosan C, Difede J, et al. Predictors of responses to psychotherapy referral of WTC utility disaster workers. *J Trauma Stress* 2006; 19(2): 307–312.
34. Covell NH, Essock SM, Felton CJ, Donahue SA. Characteristics of Project Liberty clients that predicted referrals to intensive mental health services. *Psychiatr Serv* 2006; 57(9): 1313–1315.
35. Elhai JD, Jacobs GA, Kashdan TB, et al. Mental health service use among American Red Cross disaster workers responding to the September 11, 2001 U.S. terrorist attacks. *Psychiatry Res* 2006; 143(1): 29–34.
36. Frans Ö, Rimmö P-A, Åberg L, Fredrikson M. Trauma exposure and post-traumatic stress disorder in the general population. *Acta Psychiatr Scand* 2005; 111(4): 291–299.
37. Kessler RC, Sonnega A, Bromet E, et al. Posttraumatic stress disorder in the National Comorbidity Survey. *Arch Gen Psychiatry* 1995; 52: 1048–1060.
38. Herman D, Felton C, Susser E. Mental health needs in New York State following the September 11th attacks. *J Urban Health* 2002; 79(3): 322–331.
39. Hasin DS, Goodwin RD, Stinson. Epidemiology of major depressive disorder: results from the national epidemiologic survey on alcoholism and related conditions. *Arch Gen Psychiatry* 2005; 62: 1097–1106.
40. Katz CL, Smith R, Silverton M, et al. A mental health program for ground zero rescue and recovery workers: cases and observations. *Psychiatr Serv* 2006; 57(9): 1335–1338.
41. Johnson SB, Langlieb AM, Teret SP, et al. Rethinking first response: effects of the clean up and recovery effort on workers at the world trade center disaster site. *J Occup Environ Med* 2005; 47(4): 386–391.
42. van der Kolk BA, van der Hart O. Pierre Janet and the breakdown of adaptation in psychological trauma. *Am J Psychiatry* 1989; 146: 1530–1540.

Impact of September 11 World Trade Center Disaster on Children and Pregnant Women

Philip J. Landrigan, MD, MSc,[1] Joel Forman, MD,[2] Maida Galvez, MD, MPH,[1]
Brooke Newman, BA,[1] Stephanie M. Engel, PhD[1] and Claude Chemtob, PhD[3]

[1] Department of Community and Preventive Medicine, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, NY, USA
[2] Department of Pediatrics, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, NY, USA
[3] Department of Psychiatry, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, NY, USA

ABSTRACT

Background: Children are uniquely sensitive to toxic exposures in the environment. This sensitivity reflects children's disproportionately heavy exposures coupled with the biologic vulnerability that is a consequence of their passage through the complex transitions of early development.

Methods and Results: To assess effects on children's health associated with the attacks on the World Trade Center (WTC) of September 11, 2001, research teams at the Mount Sinai School of Medicine and other academic health centers in New York City launched a series of clinical and epidemiologic studies. Mount Sinai investigators undertook a prospective analysis of pregnancy outcomes in 182 women who were pregnant on September 11, 2001, and who had been either inside or within 0.5 miles of the WTC at the time of the attacks; they found a doubling in incidence of intrauterine growth retardation (IUGR) among infants born to exposed mothers as compared to infants born to unexposed women in northern Manhattan. A Columbia research team examined pregnancy outcomes in 329 women who lived, worked or gave birth in lower Manhattan in the 9 months after September 11; they found that these women gave birth to infants with significantly lower birth weight and shorter length than women living at greater distances from Ground Zero. NYU investigators documented increased numbers of new asthma cases and aggravations of preexisting asthma in children living in lower Manhattan. Mount Sinai mental health researchers documented a significant increase in mental health problems in children who directly witnessed the attacks and subsequent traumatic events; these problems were most severe in children with a past history of psychological trauma. The New York City Department of Health and Mental Hygiene established a WTC Registry that has enrolled over 70,000 persons of all ages in lower Manhattan and will follow the health of these populations to document on a continuing basis the health consequences of September 11. ***Mt Sinai J Med 75:129–134, 2008.*** **©** *2008 Mount Sinai School of Medicine*

Key Words: World Trade Center, September 11, pregnant women, children, environmental exposures.

The terrorist attacks on the World Trade Center (WTC) of September 11, 2001 and the subsequent collapse of the twin towers created an acute environmental disaster of unprecedented magnitude in New York City. The combustion of more than 90,000 L of jet fuel at temperatures above 1,000°C released a dense and intensely toxic atmospheric plume containing soot, metals, volatile organic compounds (VOCs), and hydrochloric acid. The collapse of the buildings pulverized cement, glass, and building contents and generated thousands of tons of particulate matter (PM) composed of cement dust, glass fibers, asbestos, lead, polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), organochlorine pesticides, and polychlorinated dioxins and furans.[1–3] These materials dispersed over

Address Correspondence to:

Philip J. Landrigan
Mount Sinai School of Medicine
Department of Community and
Preventive Medicine
New York, NY, USA
Email: phil.landrigan@mssm.edu

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20032

© 2008 Mount Sinai School of Medicine

lower Manhattan, Brooklyn and for miles beyond. They entered nearby offices, schools, and residential buildings. Much remained at the site to form Ground Zero, a six-story pile of smoking rubble that burned intermittently for more than 3 months.[4]

Workers at Ground Zero and the public in lower Manhattan, including pregnant workers, infants and children, were exposed to smoke and dust in ambient air, in their communities, in their schools, and in their homes. The release by government authorities of inaccurate and misleading information on air quality in lower Manhattan may have aggravated the health impacts of these airborne exposures by creating a false sense of security that deterred appropriate use of respiratory protection.

In the aftermath of September 11, the National Institute of Environmental Health Sciences (NIEHS) gathered together a consortium of academic research centers in environmental health for the purpose of organizing studies to examine the environmental consequences and human health impacts of the destruction of the WTC. These studies included chemical analyses of the dust that settled at Ground Zero, air monitoring, study of the plumes of smoke and dust released to the atmosphere, and health studies of exposed populations. The populations of greatest concern were workers at Ground Zero, such as police, firefighters, construction workers and building cleaners, and vulnerable populations living, working and attending school in lower Manhattan, especially pregnant women and young children.

PREGNANCY OUTCOME STUDIES

To investigate health impacts in the vulnerable populations in lower Manhattan, researchers at the Mount Sinai School of Medicine and the Columbia University Center for Children's Health developed conjoint proposals to examine infants born to women who were pregnant on September 11, 2001 and who were either acutely or chronically exposed to smoke and dust from the twin towers. In the Mount Sinai study, the acute exposure group consisted of pregnant women who were actually in the WTC towers or in nearby office buildings at the time of the attacks. The chronic exposure group consisted of women who lived, worked or delivered their babies in the communities of lower Manhattan in the weeks and months after September 11; this group has been followed by researchers from Columbia.

To assess the exposures of these women and their infants to toxic chemicals released into the environment by the destruction of the towers, samples of blood and other biological fluids were taken during pregnancy or at delivery. To further characterize exposures, efforts were made to reconstruct the women's locations in the aftermath of September 11th and then to correlate that information with geocoded data on environmental exposure levels.[5] To assess infants' growth, development and health status, each infant was evaluated at birth and periodically reexamined over the first several years of life. Concomitant psychological examinations of mothers and their babies have also been conducted.

Mount Sinai Pregnancy Outcome Study

To evaluate whether acute, high-level exposures to environmental toxins on and after September 11, were related to impaired fetal growth or other adverse pregnancy outcomes, researchers from the Mount Sinai School of Medicine established a prospective epidemiologic study of 187 women who were pregnant and either within or near the WTC on or immediately after the attacks; 12 of these women were actually in the towers, and nearly all of the rest were within one-half mile.[6] All of these women were identified through their obstetricians, and they all provided fully informed, written consent to participate in the study.

These highly exposed women experienced no increased frequency of miscarriages and no increases in incidence of either premature or low-birth weight infants. However, a striking finding was that these acutely exposed women experienced a twofold-increased risk of delivering small-for-gestational age (SGA) infants. SGA infants are defined as infants with a birth weight below the 10th percentile for gestational age based on the nomogram by Brenner.[7] This rate of SGA was significantly elevated compared to a population of more than 3,000 women of similar ethnic and socioeconomic background who delivered at Mount Sinai Hospital during the same 9-month time period, and had no direct exposure to dust and smoke from the WTC.

Examination of potential confounders such as marital status, education, pre-pregnancy weight, and pregnancy-induced hypertension did not materially alter the results. No significant difference in the frequency of SGA was observed according to the trimester of pregnancy at the time of the attacks.

This finding of an increased incidence of SGA infants among women acutely exposed suggests that the WTC disaster had a detrimental impact on their pregnancies. The Mount Sinai investigators speculated that women's exposures to smoke and dust from the towers may have contributed to this effect, and they noted that an increased incidence of SGA is seen among women who smoke during pregnancy; intrauterine exposure to products of combustion is thought to account for that effect.[6]

In addition to toxic chemical exposures, the destruction of the WTC was a source also of enormous psychological trauma. To assess the potential impact of this psychological trauma on pregnant women and their fetuses, the Mount Sinai investigators undertook mental health assessments of women with singleton pregnancies included in the study described above. A total of 52 women completed at least one psychological assessment prior to delivery. In adjusted multivariable models, both post-traumatic stress symptomatology (PTSS) and moderate depression were associated with longer gestational durations. Further it was found that PTSS was associated with decrements in infant head circumference at birth ($\beta = -0.07$, SE = 0.03, $p = 0.01$). The investigators opined that the impact of stress resulting from extreme trauma may be different from that resulting from ordinary life experiences, particularly with respect to cortisol production. The observed decrements in head circumference may influence subsequent risk of neurocognitive development. Long-term follow-up of these infants is needed to evaluate the persistence of these effects.[8]

Columbia Pregnancy Outcome Study

The Columbia Center for Children's Environmental Health launched a study in December 2001 to examine effects on gestational age, birth size, respiratory health and neurocognitive development of chronic exposure during pregnancy to WTC contaminants. This study was undertaken in a group of 329 women who lived, worked or gave birth in lower Manhattan during the months after September 11.

Of 738 women initially screened for eligibility, 369 women were eligible and gave their consent for participation; 329 contributed at least one blood sample (cord or maternal blood), medical record information, and a complete postpartum interview, all of which were required for full enrollment in the study.[9] All provided fully informed written consent to participate in the study.

The researchers found that the participating women who lived within two miles of the WTC delivered infants with significantly lower weight (−149 g) and reduced length (−0.8 cm) at birth as compared to women who lived at greater distances. Only part of this observed reduction could be accounted for by shortening of duration of gestation. Mothers who were in their first trimester of pregnancy on September 11 delivered infants with statistically significantly shorter gestation duration (−3.6 days) and smaller head circumference (−0.48 cm), compared to women who were in later stages of pregnancy on that date, regardless of the distance of their residences or work sites from the WTC.[10]

Further work in the Columbia Population has documented associations between biomarkers of exposure and proximity to the site and has also found associations between exposure biomarkers and developmental outcomes (http://www.CCCEH.org).

MOUNT SINAI PEDIATRIC HEALTH STUDY

To examine the behavioral consequences in preschool children of exposure to the WTC attacks, Mount Sinai investigators conducted a retrospective cohort study of children living in lower Manhattan.[11] A total of 116 preschool children directly exposed to the WTC attacks participated in this study. The psychological traumas assessed included seeing people jumping out of the towers, seeing dead bodies, seeing injured people, and witnessing the towers collapsing. Past history of exposure to psychological trauma was also assessed. A child was considered exposed to psychological trauma following September 11, if the child had experienced at least one of the above listed events. The main outcome measure was clinically significant behavioral problems as measured by the Child Behavioral Checklist.

Preschool children exposed to high-intensity WTC attack-related events were found to be at increased risk for sleep problems and for anxious/depressed behavioral symptoms. Children who were directly exposed to high-intensity WTC attack-related events and who had a past history of psychological trauma were especially likely to have clinically significant emotionally reactive, anxious/depressed, and sleep-related behavioral problems.[11]

MOUNT SINAI PEDIATRIC OUTREACH

Mount Sinai's Pediatric Outreach Program worked extensively after September 11th in lower Manhattan with residents, businesses, schools, churches and community organizations. The goal of this outreach effort was to provide science-based, objective information to parents, pediatricians, government officials and caretakers (i.e. nursery school, day-care and school personnel) about potential environmental threats to children's health. Pediatricians in the program answered questions, formulated plans for clean up and for prevention of further exposures, served as an unbiased independent source of analysis, and disseminated information to a wide array of audiences.

The four postal zip codes closest to Ground Zero contain 36 schools serving approximately 19,000 children. Thousands of additional preschoolers reside and/or attend day care and nursery schools in lower Manhattan. The Mount Sinai efforts were aimed at coordinating outreach and education in these schools and nurseries to reduce the very heightened level of anxiety among parents, school officials, area residents and other individuals. Thousands of people were reached through these efforts. The TV interview with Bill Moyers on October 19, 2001 was especially important.

The Pediatric Outreach Program collaborated with the WTC Worker and Volunteer Medical Screening Program at Mount Sinai to provide pediatric environmental consultation to adolescents who had worked at Ground Zero. The Program continues also to answer calls and respond to media reports. After a lull in public focus on the environmental aftermath of September 11, there is now a new wave of concern around the planned demolition of the Deutsche Bank Building and the reconstruction of lower Manhattan. The Program stands ready to respond to calls from members of the community.

To further educate pediatricians about environmental threats to children's health, the Pediatric Outreach Program piloted an educational seminar on November 15, 2004 titled 'The World Trade Center Disaster: Effective Case Management of Environmentally Exposed Children' at the New York Academy of Medicine. The audience consisted of a select group of practitioners and government representatives. The Program utilized the WTC experience as a case study in pediatric environmental exposure to launch a training session focused on efficient evaluation of environmentally exposed children. Topics included a review of the exposures encountered after September 11, Internet and consulting resources available to the practitioner to evaluate environmentally exposed children, and risk communication techniques. The seminar was very well received and the feedback has led to a number of modifications and enhancements.

NEW YORK UNIVERSITY HEALTH ASSESSMENT OF COMMUNITY RESIDENTS

To assess prevalence of new-onset respiratory symptoms after September 11, 2001 among previously healthy persons in lower Manhattan as well as in residents with preexisting asthma, a team from New York University Medical Center in collaboration with the New York State Department of Health and the New York Academy of Medicine conducted a clinical and epidemiologic survey.[12] Symptoms were assessed by questionnaire, and pulmonary function was evaluated in a subset of the study population by standard screening spirometry.

A total of 2,166 residents of lower Manhattan living within a 1.6-km radius of the WTC were enrolled in this survey and compared with 200 persons living 1.6–8.0 km distant. Spirometry was performed in 52 residents. Preliminary data indicate that previously healthy persons living near Ground Zero had a greater increase in prevalence of respiratory symptoms after September 11 than did more distant residents. These symptoms were predominantly cough, wheeze, and shortness of breath. Symptoms were not associated with abnormal screening spirometry.

Preexisting asthmatic residents in the exposed area also reported a higher prevalence of respiratory symptoms after September 11. They also reported an increased use of asthma medication relative to controls. These findings are consistent with observations in exposed workers and also with findings observed in laboratory animals exposed experimentally to WTC dust.[13]

NEW YORK CITY DOHMH HEALTH REGISTRY

In 2004, to facilitate continuing follow-up of the populations exposed in the aftermath of September 11, the New York City Department of Health and Mental Hygiene (NYC DOHMH), together with the Center for Disease Control's Agency for Toxic Substances and Disease Registry (ATSDR), established the World Trade Center Health Registry (WTCHR). To establish this Registry, baseline interviews were conducted on more than 71,000 highly affected eligible enrollees.[14–16]

The WTCHR is a vital resource that will enable public health and medical experts to systematically monitor the health of tens of thousands of persons potentially impacted by the September 11, 2001 attacks on the WTC. The WTCHR will make it possible for those who enrolled in the Registry and thousands of others who may have been exposed to receive the information and advice they want and need to make informed decisions about their health. In order for the WTCHR to fulfill its potential as a public health resource, an ongoing effort is needed to maintain contact with WTCHR participants and sustain their interest so that they can continue to provide information to the registry and receive feedback on what is learned about the health impact of the September 11 attacks. DOHMH will maintain contact and follow all Registry enrollees for up to 20 years to determine health effects resulting from the environmental and mental health exposure to the attack. It has already documented a two-fold increase in new onset asthma in exposed children[17–18].

CONCLUSIONS

The data that have emerged from lower Manhattan in the aftermath of September 11 indicate that health effects occurred in vulnerable populations, but that these effects were in general less severe and qualitatively different from the effects observed in the heavily exposed worker populations.

Two major lessons emerge from the WTC experience that should inform preparation and response to future disasters, whether natural or manmade. The first of these lessons is that preparedness is critical. There need to be prepositioned resources in place before the occurrence of a disaster to deal with the inevitable impacts on both the physical health and the mental health of exposed populations, especially vulnerable subgroups. These resources must include human resources–trained personnel and robust organizations–as well as fiscal resources, such as contingency funds in the budgets of CDC, NIH and the Department of Homeland Security that can be rapidly and effectively deployed to qualified organizations in the aftermath of a disaster.

A second lesson is that long-term, continuing follow-up will be needed to fully assess and treat disaster-related health effects. Thus, long-term follow of the infants exposed *in utero*, and follow-up through the World Trade Center Registry of the general population who resided in lower Manhattan in the days and weeks after September 11 will be required to characterize and manage the long-term and delayed consequences of the disaster and thus to chart the full magnitude of the health impact of the attacks on the WTC.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support for this work that was provided by the The Mount Sinai Superfund Basic Research Program (P42 ES007384) and The Mount Sinai Center for Children's Environmental Health and Disease Prevention Research (P01 ES009584 and EPA RD 83171101) and the New York Community trust. We also acknowledge the support of Drs. Mary Wolff and Gertrud Beckowitz of the Mount Sinai School of Medicine, Dr. Frederic P. Pererera of the Mailman School of Public Health of Columbia University and of Dr. Lorna Thorpe of the New York City Department of Health and Mental Hygiene.

REFERENCES

1. Clark RN, Green RO, Swayze GA, et al. Environmental Studies of the World Trade Center Area after the September 11, 2001 Attack, 2003; Available: http://pubs.usgs.gov/of/2001/ofr-01-0429/ [accessed 16 September 2005].
2. Lioy PJ, Weisel CP, Millette JR, et al. Characterization of the dust/smoke aerosol that settled east of the World Trade Center (WTC) in Lower Manhattan after the collapse of the WTC 11 September 2001. *Environ Health Perspect* 2002; 110: 703–714.
3. McGee JK, Chen LC, Cohen MD, et al. Chemical analysis of World Trade Center fine particulate matter for use in toxicological assessment. *Environ Health Perspect* 2003; 111: 972–980.
4. Landrigan PJ, Lioy PJ, Thurston G, et al, NIEHS World Trade Center Working Group. Health and environmental consequences of the world trade center disaster. *Environ Health Perspect* 2004; 112(6): 731–9, Review.
5. Wolff MS, Teitelbaum SL, Lioy PJ, et al. Exposures among pregnant women near the World Trade Center site on 11 September 2001. *Environ Health Perspect* 2005; 113(6): 739–48.
6. Berkowitz GS, Wolff MS, Janevic TM, et al. The World Trade Center disaster and intrauterine growth restriction. *JAMA* 2003; 290(5): 595–596, No abstract available. Erratum in: JAMA. 2003 Dec 10; 290(22): 2943.
7. Brenner WE, Edelman DA, Hendricks CH. A standard of fetal growth for the United States of America. *Am J Obstet Gynecol* 1976; 126: 555–564.
8. Engel SM, Berkowitz GS, Wolff MS, Yehuda R. Psychological trauma associated with the World Trade Center attacks and its effect on pregnancy outcome. *Paediatr Perinat Epidemiol* 2005; 19(5): 334–341.
9. Perera FP, Tang D, Rauh V, et al. Relationships among polycyclic aromatic hydrocarbon-DNA adducts, proximity to the World Trade Center, and effects on fetal growth. *Environ Health Perspect* 2005; 113(8): 1062–1067.
10. Lederman SA, Rauh V, Weiss L, et al. The effects of the World Trade Center event on birth outcomes among term deliveries at three lower Manhattan hospitals. *Environ Health Perspect* 2004; 112(17): 1772–1778.

11. Chemtob CM, Nomura Y, Abramovitz RA. Impact of conjoined exposure to the world trade center attacks and to other traumatic events on the behavioral problems of preschool children. *Arch Pediatr Adolesc Med* 2008; 162(2): 126–133.
12. Reibman J, Lin S, Hwang SA, et al. The World Trade Center residents' respiratory health study: new-onset respiratory symptoms and pulmonary function. *Environ Health Perspect* 2005; 113(4): 406–411.
13. Chen LC, Thurston G. World Trade Center cough. *Lancet* 2002; 360(suppl): S37–S38.
14. Brackbill RM, Thorpe L, DiGrande L, et al. Surveillance for World Trade Center Disaster Health Effects. *Morbidity and Mortality Weekly Report*. *55* (No. SS-2) April 2006.
15. Perrin MA, DiGrande L, Wheeler K, et al. Difference in PTSD prevalence and associated risk factors among World Trade Center rescue and recovery workers. *American Journal of Psychiatry* 2007; 164: 1385–1394.
16. Wheeler K, McKelvey W, Thorpe L, et al. Asthma Diagnosed after September 11, 2001 among rescue and recovery workers. *Environmental Health Perspectives* 2007; 115: 1584–1590.
17. DiGrande L, Perrin M, Thorpe I, et al. Posttraumatic stress symptoms, PTSD, and risk factors among Lower Manhattan residents 2–3 years after the September 11, 2001 terrorist attacks. *J Traumatic Stress, in press.*
18. Thomas P, Brackbill R, Thalji L, et al. Children exposed to the World Trade Center disaster of September 11, 2001. Respiratory and psychological health. *Env Health Perspectives, in press.*

Responder Safety and Health: Preparing for Future Disasters

Dori B. Reissman, MD, MPH and John Howard, MD, JD

National Institute for Occupational Safety and Health, Washington, D.C., USA

ABSTRACT

This article reviews lessons learned about managing the safety and health of workers who were involved in disaster response, recovery, and cleanup after the 2001 World Trade Center (WTC) disaster. The first two sections review ongoing responder health burdens and the tragic toll of this disaster from a worker safety and health perspective. The remaining sections address changes in federal infrastructure, response planning, and resources for protection of response and recovery personnel.

Proper preparation includes pre-event and "just-in-time" disaster-worker training on likely hazards, organizational assets for hazard monitoring, and hands-on instruction in the use of assigned protective equipment. Good planning includes predeployment medical review to ensure "fitness for duty" and considers the following: (1) personal risk factors, (2) hazards likely to be associated with particular field locations, and (3) risks involved with assigned tasks (eg, workload and pace, work/rest cycles, available resources, and team/supervisory dynamics). Planning also should address worker health surveillance, medical monitoring, and availability of medical care (including mental health services). Disaster safety managers should anticipate likely hazards within planning scenarios and prepare asset inventories to facilitate making timely safety decisions.

Disaster safety management begins immediately and provides ongoing real-time guidance to incident leadership at all levels of government. Robust standards must be met to reliably protect workers/responders. An integrated and measurable multiagency safety management function must be built into the incident command system before an incident occurs. This function delineates roles and responsibilities for rapid exposure assessments, ensuring cross-agency consistency in data interpretation, and timely, effective communication of information and control strategies. The ability to perform this safety mangement function should be tested and evaluated in exercise simulations and drills at multiple levels.

Joint planning and exercising of the safety management plan and its function are effective ways to build interagency relationships and to be more systemic in managing logistics for safety equipment and converging personnel. Planning must include mechanisms to enable safety decisions to be implemented–such as effective and rapid scene control (site access), personnel tracking, and safety enforcement. Worker safety and health preparedness and leadership are essential for protecting workers and promoting resiliency among personnel involved in disaster response, recovery, and cleanup. **_Mt Sinai J Med 75:135–141, 2008._** © *2008 Mount Sinai School of Medicine*

Key Words: worker safety, disaster-worker health, responder health, responder safety, disaster-safety management.

Address Correspondence to:

Dori B. Reissman
Office of the Director
National Institute for
Occupational Safety and Health
Washington, D.C., USA
Email: DReissman@cdc.gov

BACKGROUND ON THE WORLD TRADE CENTER DISASTER

Terrorists hijacked 4 commercial jet airliners that were carrying passengers and full complements of jet fuel on September 11, 2001. Each hijacked aircraft was used as an unconventional weapon, was filled with jet fuel, was moving at more than 300 mph, and was targeting major US symbols of power (Pentagon) and financial prowess (World Trade Center [WTC]). The first Boeing 767 crashed into the WTC North Tower in lower Manhattan, followed by a second airplane that crashed into the South Tower 14 minutes later. Initially, the command center for rescue operations was set up on the ground floor of the towers and on nearby streets, according to protocol for

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20024

© 2008 Mount Sinai School of Medicine

a high-rise fire in this vicinity.[1] Unfortunately, the multitude of rescuers quickly ascending the burning WTC towers to assist with evacuation/rescue and fire suppression were not aware of the imminent loss of structural integrity in either building. About 90,000 L of jet fuel was burning at high temperatures and weakening the steel framework of the towers.[2] Within 2 hours, both towers dramatically collapsed, with fires subsequently destroying 7 WTC (housing the Emergency Operations Center for New York City) and 6 WTC (housing the Manhattan area office of the Occupational Safety and Health Administration [OSHA]). The death toll was 2801 people–including 343 firefighters from the Fire Department of New York City (FDNY), 23 officers from the New York City Police Department (NYPD), 37 officers from the Port Authority of New York and New Jersey, 2251 building occupants, and 127 aircraft passengers and 20 crew members.[3]

Thousands of responders and volunteers quickly converged to assist with rescue operations, with a continual influx of many thousands trained and spontaneous (unaffiliated) volunteers arriving over the ensuing days and months to help with the sustained response campaign. The WTC site continued to burn intensely until early October 2001, forming a 6-story pile of rubble called "Ground Zero." Demolition and removal of thousands of tons of debris then began while fires continued to smolder and release toxic combustion products.[4] The last debris fire was extinguished on December 20, 2001. More is known about the content of settled dust than about suspended particulates and gases, especially during the first 24 hours (most intense plume) when testing was not conducted.[5] Much of the settled dust was composed of construction materials such as pulverized cement, wallboard, office furnishings, and glass fibers. However, the coarse and fine particulates were found to be very alkaline in nature (pH ranging from 8 to >11) and, thus, likely to be corrosive to the respiratory tract lining.[6] Numerous hazardous substances were identified within both settled and airborne dust, including asbestos, silica, and polyaromatic hydrocarbons (PAH)–agents that may cause cancer or pneumoconiosis over the long term.

Attempts to set up screening and medical monitoring programs for WTC responders have been administratively challenging for a number of reasons–including the lack of adequate systems for personnel tracking (knowing who was there, when, and for how long), contractual and financing issues, and the intricacies of standardizing clinical and data collection and analysis procedures.[7] There were acute medical problems among early responders–such as reactive airway symptoms from inhaling smoke and toxic dust, injuries from a multitude of physical hazards associated with the rubble pile, and acute stress reactions from the traumatic nature of these events, and the loss of responder comrades.[8] Given the significant increase in persistent respiratory complaints detected early among FDNY responders, medical screening programs for other responders were initiated.[9] Philanthropic organizations were the first to provide funding for medical screening of WTC responders. In 2002, congress provided $12 million to enable other responders to receive medical screening and to allow researchers to standardize data collection for epidemiological study. The WTC Worker and Volunteer Medical Screening Program (MSP) was created as a multicenter clinical program that provided a once-only free standardized screening examination, including assessment of exposure, medical, and psychiatric complaints, and provided occupational health education and service referral.[10] Testing included a physical examination, pre- and postbronchodilator spirometry, blood cell counts and chemistry, urinalysis, and chest radiography. Eligibility was determined by timing (when responders arrived) and exposure duration (how many hours they worked). FDNY and State of New York employees were not eligible, as they had access to other programs. Early findings of adverse health effects among responders prompted the appropriation of an additional $90 million in 2003 to extend the MSP and to add a 5-year medical monitoring component, with repeat examinations every 18 months. A cross-sectional study of firefighters in New York City conducted in early October 2001 found that 19% of participating firefighters did not use a respirator, and 50% reported using a respirator only rarely.[9] Banauch and colleagues examined changes in firefighter lung function by using a retrospective longitudinal study of 12,079 rescue personnel from FDNY, 11,766 of whom had some sort of exposure to the WTC site.[11] By using mixed linear random-effect modeling, Banauch analyzed the difference in average spirometric measurements from the 5 years before September 11, 2001 to the first year after September 11, 2001 and whether WTC exposure intensity influenced spirometric changes. Exposure intensity was determined by the time of arrival, with the referent group arriving at the WTC site on or after September 13, 2001. The demographics of FDNY personnel reporting no exposure to the WTC site were too disparate to use for comparison. The analysis revealed an average-adjusted decline equivalent to 12 years of normal age-related decreases in lung function (average adjusted-FEV_1 loss of 372 mL, 95% confidence interval [CI] 364–381, $P < .001$). The decrement in lung function correlated linearly

($P = .048$) with increased exposure, as estimated by time of arrival, and respiratory symptoms predicted a further FEV_1 decrease ($P < .001$).

In 2006, the MSP reported that 69% of 9442 study responders experienced new or worsened respiratory symptoms, with persistence in 59% of these workers.[10] The most prevalent industrial sectors in the cohort were technical and public utility, law enforcement, and construction, with at least 80% belonging to a labor union. Almost one-third of all responders in the MSP had persistently abnormal spirometry–most commonly low-forced vital capacity (FVC)–with a higher likelihood of such problems noticed among responders who arrived closer to the time when the twin towers collapsed. In a recent analysis of WTC Health Registry data, 926 workers (3.6% of study sample) reported newly diagnosed asthma, which, again, appeared associated with early arrival (dust-cloud exposure), longer duration of work, and longer delays in initial use of masks or repirators ($P < .001$).[12] Various studies have reported considerably different rates of psychopathology among rescue and response personnel, at least in part because of the use of various assessment methods and study eligibility criteria. In analyzing rescue/recovery worker data from the voluntary WTC Health Registry, Perrin and colleagues noted that the overall prevalence of post-traumatic stress disorder (PTSD) among 28,962 worker participants was 12.4%, ranging from 6.2% for police to 21.2% for unaffiliated volunteers.[13] After adjusting for demographics, direct traumatic exposure (ie, evacuated the WTC, witnessed trauma, exposed to the dust cloud, and/or injured), and type of work at Ground Zero (ie, operated heavy equipment, performed light construction, welded, firefought, hand dug, and/or searched and rescueed), unaffiliated volunteers (adjusted odds ratio [OR_{adj}] = 3.7; 95% CI: 3.1–4.4) and construction/engineering workers ($OR_{adj} = 3.8$; 95% CI: 3.2–4.6) had the highest prevalence of PTSD. Sanitation workers ($OR_{adj} = 2.7$; 95% CI: 2.1–3.4) and individuals affiliated with volunteer organizations ($OR_{adj} = 2.0$; 95% CI: 1.6–2.5) had the next highest PTSD prevalence. The MSP found that 6.4% of the 10,133 participating WTC workers examined between 10 and 61 months after the collapse of the WTC reported symptoms meeting the criteria for probable PTSD.[14] Other psychiatric morbidities were present including probable depression reported in 8.8%, probable panic disorder in 5.0%, and probable generalized anxiety disorder in 0.9%. Odds ratios for a WTC-responder with probable PTSD also to be diagnosed with probable depression (OR = 17, 95% CI: 1.9–2.9) or panic disorder (OR = 6.9, 95% CI: 5.5–8.5) were significant and high.

Given the magnitude of health burdens emerging in this group, $75 million was appropriated in 2006 to provide services to meet the healthcare needs of MSP-participating responders who developed chronic conditions related to their WTC response work–now termed the WTC Medical Monitoring and Treatment Program.[15] Numerous legal battles over financial and legal accountability have ensued because of shortfalls in coverage arising from inadequate or absent personal insurance and workman's or victim's compensation policies–which have been especially difficult for volunteers, immigrant day laborers, or responders who were working outside their normal jurisdictions.[16] Projected long-term costs for ongoing medical monitoring and treatment in New York have ranged from $238 million to $392 million per year.[7,17]

Unfortunately, our understanding of responder safety and health from the WTC disaster is severely constrained by insufficient strategies during and immediately after the event to 1) assess and characterize environmental contamination, 2) quantify individual occupational exposures, and 3) control and manage the disaster site to ensure safe-work practices and prudent use of personal protective equipment (PPE).[18,20] In December 2001, the National Institute for Occupational Safety and Health (NIOSH) convened a conference in New York City, organized and led by the RAND Science and Technology Policy Institute, to explore lessons on protecting the safety and health of emergency responders in the context of terrorist attacks. Worker-safety and health lessons from this and other large-scale and complex domestic disasters were reviewed, with key findings and recommendations published in a 4-volume series.[18–21] These published findings reveal that on-scene professional responders and managers noted that Ground Zero site management and perimeter control were initially chaotic because of the emotionally charged atmosphere and confusion over boundaries of authority, especially as key leaders were lost in the building collapses.

Emotions ran high as responders worked frantically under dire circumstances to try and find missing comrades in the rubble. There were more than 30 entry points to the WTC site initially (external perimeter), which made scene control very difficult and personnel tracking impossible. Safety and health officials reported that erratic credentialing procedures, multiple logistical challenges (supply shortages, disrupted telecommunication systems, and faulty instrument readings due to intense smoke), and the lack of unified authority thwarted efforts to enforce appropriate PPE use and safe-work practices.

Conference participants reported that safety and health personnel were not permitted early access to the disaster site, and it took a week before coordinated daily hazard information briefings were instituted at Ground Zero. At this point, the initial intensity (first phase) of the exposure had dissipated (rainfall occurred on September 14), and, subsequently, efforts to reconstruct exposure or estimate dose during this critical period were hampered. Despite implementation of daily briefings, hazard and risk information did not flow effectively to the front-line responders across all involved organizations.[18] Situations existed where different contractors working next to each other on the same task were wearing different levels of protection. Lack of coordination among multiple entities conducting hazard assessments led to confusion about responsibilities for monitoring, appropriate contaminants to assess, and clearance thresholds; some of these issues remain unresolved (ie, thresholds for community clearance; appropriate contaminants).[20] Inconsistencies in reporting and interpreting findings and safe thresholds from different federal, state, and local sources created dilemmas, distress, and mistrust in managing risk and uncertainties. Shortcomings in site information, training, and enforcement made it difficult to control internal perimeters or hazard zones set up to indicate different needs for training, PPE, and special work practices.[18,20] Subgroups of workers were not familiar with the PPE being dispensed or were unfamiliar with basic safe-work practices such as not consuming food and beverages in contaminated areas. PPE was sometimes selected on the basis of appearance or comfort rather than on knowledge of the hazards or the intended design of the technology.

The scale of this disaster, the prolonged rescue and recovery campaigns, the complexity of hazards encountered, and the need to preserve evidence (crime scene) required many responders to take on tasks (new roles) for which they were insufficiently prepared (informed, equipped, and trained).[18,20]

LESSONS LEARNED: FEDERAL PLANS FOR RESPONDER SAFETY AND HEALTH

The U.S. Department of Homeland Security (DHS) was created in March 2003 in the aftermath of the 2001 multiple acts of terrorism–including anthrax spores found within the U.S. Postal Service system in October 2001. Problems with sampling methodologies and data interpretation for the anthrax contamination arose from different philosophical approaches (eg, environmental microbiology) and disconnected efforts among multiple entities charged with measuring exposure. Even as the new DHS was being assembled, the United States was facing the threat of an emerging and deadly virus causing what came to be known as severe acute respiratory syndrome (SARS). SARS was substantively an occupational disease (healthcare sector); with the heavy burden of disease occurring among hospital workers and healthcare providers in affected countries.[22]

In an effort to address lessons learned, the Worker Safety and Health Support Annex (WSH-A) was added to the National Response Plan (NRP) to help focus on and systematize protection of response and recovery personnel.[23,24] Occupational Safety and Health Administration (OSHA) personnel, in the U.S. Department of Labor, were given lead responsibility to coordinate federal assets for worker safety and health and to provide technical support when activated under the NRP.[24] These plans were approved only months before the devastation of Hurricane Katrina ravaged nearly 93,000 m^2 in several states along the Gulf Coast in August 2005.[25] The widespread nature of multiple simultaneous catastrophies in this unprecedented storm season overwhelmed response capabilities and provided challenges outside the then current disaster operational plan. Despite the magnitude and breadth of activities that resulted in the removal of 20,000 workers from serious hazards, an extensive hazard assessment effort, and interventions covering more than 50,000 workers, numerous problems were uncovered in this initial activation of the WSH-A.[26] There were conflicts over authority (who's in charge); general lack of awareness on the role of OSHA personnel plays in a disaster; and insufficient logistics to meet the needs for on-scene worker training, personal protective equipment, worker health, and injury surveillance (federal responders), and mental health service delivery.

In further efforts to address lessons learned, DHS worked with numerous stakeholders to produce a list of essential capabilities to assist planning for communities and all levels of government in support of the National Preparedness Goal.[27] In 2006, responder safety and health became 1 of these essential (targeted) capabilities and incorporated incident safety management within the National Incident Management System (NIMS) framework and key elements of the WSH-A of the NRP.[28] However, specific safety and health operational challenges and performance measures still need to be incorporated into exercise simulations and drills at federal, state, and local levels.

LESSONS LEARNED: PREPARING FOR THE FUTURE

Since Katrina, the NRP no longer requires event-related activation but is considered to be operational at all times. However, the implementation of the Support Annexes (such as the WSH-A) still requires specific activation based on situational analysis. As such, delays in activating essential federal activities for worker safety and health could compromise early decisions for safety management. Currently, federal efforts for disaster preparedness and response are being reconfigured into a National Response Framework, but this is still under development (http://www.fema.gov/emergency/nrf/mainindex.htm). These elements only apply to federal responders and their contractors; there is no coordinated system for general industry responders (such as those responding to Katrina). Disaster safety management needs to provide ongoing real-time guidance to incident leadership at all levels of government to ensure robust standards are met to reliably protect responders. Interagency coordination and unified functional command are essential for comprehensive incident-wide safety management during complex, large-scale operations.[18,20] An integrated and scalable multiagency safety management function must be built into the incident command system before an incident occurs. This safety management function needs direct access to advise the incident commander/manager or unified command about overall workforce protection needs.

Priority and authority for incident safety management need to be determined early to facilitate disaster site access, to implement strategic planning and oversight, and to rapidly engage in hazard assessments. Information management is critical, especially when multiple entities collect and interpret data on hazards. Scalable planning considers mechanisms to rapidly coordinate, vet, and disseminate hazard data and real-time control measures. Interpretation can be aided by geographic information systems, which give visual dimensions to emerging hazardous zones. Joint planning and execution of the safety management plan and functions are effective ways to build interagency relationships. Although each response organization manages its own personnel, a larger scale and multiagency operation will need a more systemic solution to manage the logistics for safety equipment and converging personnel, disaster volunteers, and others who may flock to the scene (eg, important persons or media).

Planning for effective incident-safety management must include mechanisms to enable safety decisions to be implemented—such as effective and rapid scene control (site access) and safety enforcement. Safety and health training requirements often differ across response organizations. Proper preparation includes pre-event training on likely hazards, organizational assets for hazard monitoring, and hands-on instruction in the use of assigned protective equipment. Preparedness plans and exercises should delineate roles and responsibilities for rapid assessment of exposure, ensure cross-agency consistency in data interpretation, and effectively communicate results. Event management plans need to include mechanisms to provide on-scene training on likely hazards and accessible PPE (logistics and proper use), with real-time updates as more information and PPE become available. The on-scene training (just-in-time-training) must consider the environment where training will take place as well as target audiences (education, language, skills, predicted tasks, or responsibilities). Anticipating likely hazards within planning scenarios will help identify monitoring equipment needs and technical expertise that may reside in different response organizations. Having an inventory of such assets and a strategic plan that defines priorities for assessment will facilitate safety decisions. Prior events have highlighted the need for front-line dissemination of risk and safety information and standardization of safe-work practices and interoperable safety equipment (eg, donated equipment caches). The prolonged duration of complex, large-scale response operations raises the need for PPE ensembles that are more comfortable, flexible, allow for better temperature regulation, improve job-related communication, and remain effective against a spectrum of likely hazards.[18]

Medical, emotional, and cognitive readiness are important dimensions of workforce health protection planning.[29] Good planning includes predeployment medical review to ensure fitness for duty and considers the following: 1) personal risk factors, 2) hazards likely to be associated with particular field locations, and 3) the risks involved with assigned tasks (eg, workload and pace, work/rest cycles, available resources, and team/supervisory dynamics). Personal-risk factors may include poorly controlled chronic illness (eg, hypertension, coronary artery disease, and diabetes), disability, needs for refrigerated medication or durable medical equipment, psychological flexibility and stress tolerance (coping), possession of requisite skills/training (including incident command and safety management), physical fitness, or other factors that could impact the well-being of workers and mission success. Such considerations are especially critical in chaotic, extreme, or austere environments where outside help may not be immediately available or access to water/food or electrical power is limited. Logistical preparations must include information management systems and reliable communication channels to facilitate decisions for real-time hazard control–especially given the dynamic nature of disaster scenes and issues surrounding compliance with safety decisions. Data will be needed on hazard assessments, who (personnel) is on scene, and on worker illness and/or injury. As has been described, preparedness and leadership are essential for protecting workers and promoting resiliency among personnel who are involved in disaster response, recovery, and cleanup.

OSHA electronically published information pertaining to safety and health for emergency response (http://www.osha.gov/SLTC/emergencypreparedness/index.html) and planning tools for hurricane response and recovery (http://www.osha.gov/OshDoc/hurricaneRecovery.html), with expert input from the NRP Worker Safety and Health Support Coordination Committee. NIOSH has also developed several resources and tools to assist emergency preparedness and response, including the following: 1) general medical review guidelines for disaster workers (predeployment at http://www.cdc.gov/niosh/topics/flood/preexposure.html; postexposure at http://www.cdc.gov/niosh/topics/flood/MedScreenWork.html), 2) compiled portfolio (research and products) on worker emergency preparedness (http://www.cdc.gov/niosh/programs/epr/), 3) emergency response resources (http://www.cdc.gov/niosh/topics/emres/), and 4) a compiled portfolio on exposure assessment (http://www.cdc.gov/niosh/programs/expa/).

FINAL REMARKS

As increased emphasis has been placed on the speed of disaster recovery and cleanup efforts, there must be a proportionate effort to protect workers from associated hazards. Incident-wide safety management must be ramped up quickly, implementing scalable and coordinated plans for hazard assessment, data sharing, and real-time and consistent hazard control and communication strategies. Ongoing preparedness efforts need to implement joint planning and exercises that embrace practical and rational strategies for worker/responder safety and health. Integrated disaster safety management makes good business and political sense.

DISCLAIMER

The findings and conclusions of this report are those of the authors and do not necessarily represent the views of the NIOSH, Centers for Disease Control and Prevention, or the U.S. Public Health Service.

REFERENCES

1. Smith D. *Report from Ground Zero: The Heroic Story of the Rescuers at the World Trade Center.* New York: Viking Penguin, 2002.
2. Ashley S. When the Twin Towers fell: One month after the attack on the World Trade Center, M.I.T. structural engineers offer their take on how and why the towers came down, 2001, Scientific American 2001 Oct 9. Available from: URL: http://www.sciam.com/article.cfm?articleID=000B7FEB-A88C-1C75-9B81809EC588EF21&sc=I100322.
3. Hirschkorn P. 9/11 victims' names posted at Ground Zero. Cable News Network 2002 Sep 10. Available from: URL: http://archives.cnn.com/2002/US/09/10/ar911.wtc.names.numbers/index.html. Sept 10, 2000.
4. Landrigan P, Lioy P, Thurston G, et al, NIEHS (National Institute of Environmental Health Sciences) Working Group. Health and environmental consequences of the World Trade Center Disaster. *Environ Health Perspect* 2004; 112(6): 731–739.
5. Lioy P, Pellizzari E, Prezant D. The World Trade Center aftermath and its effects on health: understanding and learning through human-exposure science. *Environ Sci Technol* 2006; 40(22): 6876–6885.
6. Lioy P, Weisel C, Millette J, et al. Characterization of the dust/smoke aerosol that Settled East of the World Trade Center (WTC) in Lower Manhattan after the collapse of the WTC 11 September 2001. *Environ Health Perspect* 2002; 110(7): 703–714.
7. GAO. September 11: HHS needs to ensure the availability of health screening and monitoring for all responders. U.S. Government Accountability Office. Report to Congressional Committees. GAO-07-892, 2007,; Available from: URL: http://www.gao.gov/new.items/d07892.pdf
8. Berrios-Torres S, Greenko J, Phillips M, et al. World Trade Center rescue worker injury and illness surveillance, New York, 2001. *Am J Prev Med* 2003; 25(2): 79–87.
9. Feldman D, Baron S, Bernard B, et al. Symptoms, respirator use, and pulmonary function changes among New York City firefighters responding to the World Trade Center Disaster. *Chest* 2004; 125(4): 1256–1264.
10. Herbert R, Moline J, Skloot G, et al. The World Trade Center Disaster and the health of workers: five-year assessment of a unique medical screening program. *Environ Health Perspect* 2006; 114(12): 1853–1858.
11. Banauch G, Hall C, Weiden M, et al. Pulmonary function after exposure to the World Trade Center in the New York City Fire Department. *Am J Respir Crit Care Med* 2006; 174(3): 312–319, Epub 2006.
12. Wheeler K, McKelvey W, Thorpe L, et al. Asthma Diagnosed after September 11, 2001 Among Rescue and Recovery Workers: Findings from the World Trade Center Health Registry. *Environ Health Perspec* 2007;

115(11): 1584–1590. Epub 2007 Aug 27. Available from: URL: http://www.ehponline.org/members/2007/10248/10248.pdf

13. Perrin MA, DiGrande L, Wheeler K, et al. Differences in PTSD prevalence and associated risk factors among World Trade Center Disaster rescue and recovery workers. *Am J Psychiatry* 2007; 164(9): 1385–1394.
14. Stellman JM, Smith RP, Katz CL, et al. Mental health morbidity and social function impairment in rescue, recovery and cleanup workers at the World Trade Center Disaster, 2007, Submitted.
15. NIOSH. The WTC medical monitoring and treatment program: a unique NIOSH funded program for the heroes of 9/11. 2007.; Available from: URL: http://www.cdc.gov/niosh/docs/2007-109/pdfs/2007-109.pdf
16. Bernott Maginnis J. The 9/11 victim compensation fund: overview and comment, The Federalist Society, 2002; Mar 27. Available from URL: http://www.fed-soc.org/publications/PubID.136/pub_detail.asp.
17. Gibbs L, Skyler E. World Trade Center Panel: Addressing the health impacts of 9/11: Report and recommendations to Mayor Michael R. Bloomberg, Feb 2007,; Available from URL: http://www.nyc.gov/html/om/pdf/911_health_impacts_report.pdf.
18. Jackson BA, Peterson DJ, Bartis JT, et al. Protecting emergency responders: Lessons learned from terrorist attacks, Santa Monica, CA, RAND Corporation CF-176-OSTP, 2002, Available at http://www.rand.org/pubs/conf_proceedings/CF176/.
19. LaTourrette T, Peterson DJ, Bartis JT, et al. Protecting emergency responders, Volume 2: Community views of safety and health risks and personal protection needs, Santa Monica, CA, RAND Corporation MR-1646-NIOSH, 2003, Available from: URL: http://www.rand.org/pubs/monograph_reports/MR1646/.
20. NIOSH/RAND, Jackson BA, Baker JC, Ridgely MS, et al. *Protecting Emergency Responders, Volume 3: Safety Management in Disaster and Terrorism Response*: National Institute for Occupational Safety and Health, Pub No. 2004-144, 2004, Available from: URL: http://www.cdc.gov/niosh/docs/2004-144/.
21. Willis HH, Castle NG, Sloss EM, Bartis JT. Protecting emergency responders, Volume 4: Personal protective equipment guidelines for structural collapse events, Santa Monica, CA, Rand Corporation: MG-425-NIOSH, 2006, Available at http://www.rand.org/pubs/monographs/MG425/.
22. The Independent SARS Commission. Campbell A (Justice). The SARS Commission: Executive summary: Spring of Fear, Volume 1, 2006 Dec, Available at http://www.sarscommission.ca/.
23. DHS. The National Response Plan. Washington D.C.: U.S. Department of Homeland Security, 2004 Dec, Available from: URL: http://www.dhs.gov/xlibrary/assets/NRP_FullText.pdf.
24. DHS. The National Response Plan. Worker Safety and Health Support Annex. Washington D.C.: U.S. Department of Homeland Security, 2004 Dec, 335–340. Available from: URL: http://www.dhs.gov/xlibrary/assets/NRP_FullText.pdf
25. Katrina Lessons Review Group. The Federal response to Hurricane Katrina: Lessons learned. Washington, DC.: The White House, 2006 Feb, Available from: URL: http://www.whitehouse.gov/reports/katrina-lessons-learned/.
26. GAO. Disaster preparedness: Better planning would improve OHSA's efforts to protect workers' safety and health in disasters, U.S. Government Accountability Office. Report to Congressional Committees. GAO-07-193. 2007. Available from: URL: http://www.gao.gov/new.items/d07193.pdf
27. Foresman GW. (Honorable). Statement of the Honorable George W. Foresman, Under Secretary for Preparedness, U.S., Department of Homeland Security, Commission on the National Guard and Reserves. Dec 13, 2006, Available from: URL: http://www.dhs.gov/xnews/testimony/testimony_1166567952901.shtm
28. DHS. Targeted Capabilities List, version 2.0. U.S. Department of Homeland Security, Nov 11, 2005, Available from: URL: www.nwcphp.org/docs/competencies/Target_Capabilities_List_2_0.pdf.
29. Reissman DB. Disaster readiness and response: what the occupational health professional needs to know. SEAK, Inc. Proceedings of the 27th Annual National Workers' Compensation and Occupational Medicine Conference; 2007 Jul 19; Hyannis Maine, 2007.

CONTROVERSY

Making Sense of ENHANCE: Ezetimibe (Zetia) Lowers LDL Cholesterol But Doesn't Decrease Carotid Intima-media Thickness

Donald Smith, MD, MPH

Lipids and Metabolism, Zena and Michael A. Wiener Cardiovascular Institute, Mount Sinai Medical Center, New York, NY, USA

On Monday, January 14, 2008, Merck and Schering & Plough released to the press an abstract being sent to the American College of Cardiology for presentation at its annual scientific meeting in March. The study, Ezetimibe and Simvastatin in Hypercholesterolemia Enhances Atherosclerosis Regression (ENHANCE),[1] compared simvastatin 80 (Zocor 80) with simvastatin 80/ezetimibe 10 (Vytorin 10/80) by using quantitative B-mode ultrasonographic changes in carotid artery intima-media thickness (IMT) over a 2-year period. The design of the study, by a very skilled team in the Netherlands, duplicated that of an earlier study, atorvastatin versus simvastatin on Atherosclerosis Progression (ASAP)[2] that compared aggressive LDL-cholesterol lowering using atorvastatin 80 (Lipitor 80) with conventional LDL-cholesterol lowering using simvastatin 40 (Zocor 40) in a similar group of subjects. Both trials enrolled in subjects with familial heterozygous hypercholesterolemia and a very high baseline mean LDL cholesterol greater than 300 mg/dL (mean USA levels approximate 130 mg/dL).

Address Correspondence to:

Donald Smith, MD, MPH
Lipids and Metabolism
Zena and Michael A. Wiener
Cardiovascular Institute
Mount Sinai Medical Center
New York, NY
Email: donald.smith@mssm.edu

In ASAP, an LDL cholesterol of 155 mg/dL with atorvastatin 80 resulted in significant regression whereas an LDL cholesterol of 192 mg/dL with simvastatin 40 resulted in only a 2-year slowing of carotid artery IMT progression. ENHANCE, a surrogate imaging trial of LDL-lowering preventive efficacy, enrolled double the number of patients as ASAP and was expected to produce similar results, which would bolster the proposed preventive clinical efficacy of ezetimibe added to simvastatin compared with simvastatin alone. *It did not*, and the avalanche of publicity that followed became a horror for us physicians, worldwide, who were writing prescriptions for ezetimibe or simvastatin/ezetimibe for 5 million patients. If taking either of these medications or prescribing one of these, what in the world is one to do?

HOW TO INTERPRET THE ENHANCE STUDY RESULTS

The subjects of the ENHANCE study were persons with familial heterozygous hypercholesterolemia randomized to simvastatin 80 *versus* simvastatin 80/ezetimibe 10 for 2 years. The primary endpoints were changes in mean combined carotid IMT measures (6 measures of distal-wall IMT–right and left-in the common carotid artery, carotid bulb, and internal carotid artery).[1]

To interpret these results, one must first examine the data and second try to put them into the context of previous trials. Table 1 shows results provided by Merck/Schering–Plough Pharmaceuticals in their abstract on January 14, 2008, supplemented by the

A different version of this article was previously published electronically on May 23, 2008.
Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20031

© 2008 Mount Sinai School of Medicine

Table 1. *ENHANCE Study Results.*

Efficacy	Simva 80/ Ezet 10	Simva 80	p
Baseline LDL cholesterol (mg/dL)[a]	319	318	–
Reduction after 2-yr treatment (%)	58	41	<0.01
On-treatment LDL cholesterol[b]	*134*	*188*	<0.01
Baseline mean carotid IMT (mm)	0.68	0.69	NS
Change in mean carotid IMT after 2 yr (mm/2 yr)	0.0111	0.0058[c]	0.29
Cardiovascular events			
CV death	2/357	1/363	NS
Nonfatal MI	3/357	2/363	NS
Nonfatal stroke	1/357	1/363	NS
Revascularization	6/357	5/363	NS
Adverse events			
Consecutive AST ≥3 ULN	2.8	2.2	NS
CPK ≥10 ULN	1.1	2.2	NS
CPK ≥10 ULN with muscle symptoms	0.6	0.3	NS

[a] After statin washout in 80% of subjects. Thus, prior, prestudy LDL cholesterol levels unknown.

[b] Calculated by present author.

[c] Apparent difference most likely a chance effect. Patients on combination were no worse.

present author's own calculations, which are shown in italics.

For 3 days after the press release of ENHANCE, the *New York Times* and other media claimed that ". . . the plaque actually grew almost twice as fast in patients taking the combination.[3]" This interpretation was most likely fostered by Steve Nissen, MD, Chief of Cardiology at the Cleveland Clinic and by the superb coronary intravascular ultrasound investigator who has proven in elegant studies that lower LDL cholesterol levels stop progression of coronary plaques, and even lower levels produce regression. He was quoted as saying the results were "shocking," and "this is as bad a result for the drug as anybody could have feared. Millions of patients may be taking a drug that does not benefit them, raising their risk of heart attacks and exposing them to potential side effects. Patients should not be given prescriptions for Zetia unless all other cholesterol drugs have failed."[3] On *Heartwire*, a popular internet news source for cardiologists, he was quoted as saying "the data show no benefit for ezetimibe on top of simvastatin. In fact, the data on both the rate of progression of atherosclerosis and cardiovascular events are trending in the wrong direction. This is a pretty clear failure. Physicians should now stop using ezetimibe or Vytorin except as a last resort."[4] In fact, such small increases in rate of common-carotid far-wall IMT progression in other studies were not statistically significant and represented nonprogression.[5]

One of Dr. Nissen's arguments for the null result was that "statins do much more than lower LDL. They increase HDL, lower triglycerides, and reduce inflammation (as measured by ultrasensitive C-reactive protein, (usCRP)). In the absence of any demonstrable effect beyond LDL lowering, nearly one million prescriptions per week are written for ezetimibe. Is this rational?"[4] It is true that ezetimibe by itself will not produce the extra benefits provided by statins mentioned above. But the combination of ezetimibe with a statin, in addition to an approximate 14%–21% further lowering of LDL cholesterol, enhances those very extra effects produced by statins, ie., further lowering triglycerides by 11%–13%, raising HDL-cholesterol (HDL-C) an additional 0.4%–1.7%, and further lowering usCRP, a marker of inflammation, 10%–17%.[6–8] So this, in fact, was not a test of the absence of such pleiotropic effects with the use of ezetimibe alone but was a test of whether enhancing these statin-produced, extra-LDL cholesterol-lowering effects would prove beneficial. Because the study demonstrated no differential beneficial imaging effects, we feel it is more profitable to examine the peculiarities of the study that gave this null effect rather than to look for negative or hazardous effects of ezetimibe which simply to date have not been demonstrated.

This was a study in persons who were 30–75 years of age, some with stable coronary heart disease and some without. Eighty percent had been on statins and thus had been at much lower levels of LDL cholesterol for unknown periods before the washout period. Did this group have significant baseline carotid IMT abnormalities? One study showed that the risk of first myocardial infarction increases with carotid IMT ≥0.82 mm, and that a progression rate of the far-wall common carotid artery IMT of $\geq$0.034 mm/year increases the risk of future clinical events.[9] The group studied in ENHANCE had neither a baseline mean IMT this great (0.68 mm) nor a progression rate this great (0.003–0.006 mm/year). The baseline mean IMT in ENHANCE was certainly less than that studied in ASAP (0.93 mm).[2] This suggests that previous use of statins in subjects in this trial may have already decreased and stabilized carotid IMT. This phenomenon was demonstrated in the ASAP extension trial; after 2 years on atorvastatin 80 and simvastatin 40 in ASAP, a 2-year extension ensued that kept the group on atorvastatin 80 on the same drug for an additional 2 years but switched all those on simvastatin 40 to atorvastain 80. The previous increase in IMT in the second year on simvastatin 40 regressed to baseline after 1 year of switching to atorvastatin 80. However, after carotid IMT decreased for the first 2 years on atorvastatin 80, it remained constant for the next 2 years on the same dose.[10]

Could this be the reason why there was no differential effect on carotid IMT between the simvastatin/ezetimibe group with LDL cholesterol of 134 mg/dL and the simvastatin-only group with LDL cholesterol of 188 mg/dL? If the 80% of the study group previously on statins had mean LDL cholesterol levels significantly less than 200 mg/dL before the 6-week drug washout period and randomization date, then the reduction in IMT may have already occurred, and 2 years at these new levels of LDL cholesterol might not have provided enough time, or enough of a differential level in LDL cholesterol, to begin seeing a differential effect on IMT. With more time, such differential levels of LDL cholesterol might have produced differential IMT rates of progression. The carotid IMT progression rate in a previous trial in the placebo cohort was 0.031 mm/year.[11] The progression rates in both groups in this trial were one-tenth of that (0.003 mm/year), suggesting that a major stabilization of progression most likely had already occurred at the beginning of the trial and simply continued throughout the 2-year trial in both groups.

It has been shown by using quantitative coronary angiography that patients with familial heterozygous hypercholesterolemia and a very high mean baseline LDL-cholesterol level of 285 mg/dL who are taking combined bile-acid binders, niacin, and lovastatin and decreasing LDL cholesterol to only 175 mg/dL can regress coronary stenoses.[12] The stenoses, and thus plaques, of such hypercholesterolemic patients are very responsive to lowering LDL cholesterol to levels that are above average in the general population. The foam-cell accumulation in the carotid arteries, the major pathological feature of increased carotid IMT,[13] should be equally regressive at such LDL-cholesterol levels. This at least makes plausible the explanation that these unique patients may already have decreased IMT to a stable level before the study began. This is a much more plausible explanation of this null trial than (1) rejecting the notion that further LDL-cholesterol lowering is ineffective in reducing IMT thickening or reducing clinical events after more than 40 years of research consistently confirming such hypotheses, or (2) suggesting that although ezetimibe can lower LDL cholesterol, it produces some unknown toxic effect that negates its beneficial LDL-cholesterol-lowering effect.

Statins lower serum LDL cholesterol predominantly through hepatic inhibition of HMG CoA reductase, an early enzyme used in the 26-step cholesterol biosynthetic pathway. The hepatocyte has many mechanisms to maintain a steady intracellular cholesterol concentration. When statin-produced cholesterol synthesis is lowered, hepatocyte LDL receptor synthesis and transport to the cell surface is increased, which results in reduced serum LDL cholesterol. Ezetimibe, on the other hand, localizes to enterocytes of the endothelium of the small intestine. It blocks the Nieman–Pick C1-like 1 (NPC1L1) receptor necessary for cholesterol absorption from micelles in the small intestine.[14] This block in intestinal cholesterol absorption results in decreased uptake of cholesterol from the gut, decreased hepatocyte intracellular cholesterol with compensatory increased LDL receptor synthesis, and again results in reduced serum LDL cholesterol. It is hard to imagine by what mechanism a medication located predominantly in the enteroctye could negate the cardiopreventive effects of its LDL-cholesterol lowering feature. Because the average human body synthesizes 800 mg of excess cholesterol daily, which cannot be biochemically catabolized, it is eliminated from the body through the liver via bile and ultimately into feces. As this 800 mg of cholesterol passes daily through the small intestine, it will be partially absorbed. Ezetimibe, by blocking the enterocyte NPC1L1-cholesterol receptor, can thus lower LDL cholesterol even in those vegetarians whose dietary intake of cholesterol is zero.

The blocking of hepatic cholesterol synthesis with statins (HMG CoA reductase inhibitors) not only upregulates hepatocyte LDL receptor synthesis as a compensatory mechanism for stabilizing intracellular hepatocyte cholesterol concentration, but it also upregulates cholesterol absorption in the intestine. Similarly, blocking the cholesterol NPC1L1 receptor with ezetimibe in the intestine and decreasing intestinal cholesterol absorption upregulates cholesterol synthesis in the liver.[15] Hence, 2 drugs taken simultaneously such as a statin and ezetimibe, which block both cholesterol synthetic and absorptive mechanisms, work extremely well together and are very potent in reducing LDL cholesterol levels.[14] Simvastatin 10 plus ezetemibe 10 taken together are as powerful in lowering LDL cholesterol as simvastatin 80 by itself.[8] Hence, patients stopping Zetia combined with a statin or Vytorin, based on a null interpretation of this study, will have to take much higher doses of the most powerful statins to attain the same low LDL cholesterol goal. Such an augmentation of statin dose risks a small (1%–2%), but definite, increase in risk of experiencing statin side effects, such as reversible increases in serum transaminase levels or uncomfortable myalgias.

CONCLUSION AND RECOMMENDATIONS

ENHANCE is the first unfortunate imaging trial that tested whether ezetimibe added to high-dose statin could add additional benefit to carotid artery IMT thickening in persons with familial heterozygous hypercholesterolemia. Unfortunately, 80% of the subjects had been on statins for unknown periods of time, which might have already resulted in regression of IMT thickness to baseline levels less than those seen in other trials, in turn resulting in nonprogressive IMT thickening in both groups. Despite implications in the public media, the study showed no harm in adding ezetimibe to simvastatin, rather it showed only a lack of differential imaging efficacy. This is an inadequate trial of the efficacy of ezetimibe-lowered LDL cholesterol levels used to produce beneficial imaging and clinical effects. We still must wait for results in ongoing clinical efficacy trials that will be reported in 2 years.

Table 2. *National Cholesterol Education Program Adult Treatment Panel Guidelines.*

Number of risk factors[a] or medical condition	LDL cholesterol goal (mg/dL)
0, 1 risk factor	<160
Two or more risk factors	<130
Diabetes, hypertension	<100
Coronary heart diease, stroke or peripheral vascular disease	<70

[a] Risk factors to be counted:
- Hypertension history or use of antihypertensive medications
- Cigarette smoking
- Age: Men ≥45 years; women ≥55 years or postmenopausal
- Early family history of coronary heart disease: Father, brother ≤55 years; Mother, sister ≤65 years
- HDL cholesterol <40 mg/dL.

Subtract one risk factor for HDL cholesterol >60 mg/dL.

In the meantime, clinicians should focus on lowering LDL cholesterol to the proper goal, as established by the National Cholesterol Education Program (NCEP) (Table 2).

If necessary to achieve the appropriate NCEP goal comfortably, clinicians should use ezetimibe or simvastatin/ezetimibe. Indeed, we should use any and all means studied, which lower LDL cholesterol including bile-acid binders, diets, ileal bypass, and statins, have shown coronary heart-disease preventive effects. This should be true of ezetimibe as a single agent or in combination with a statin. The combination with a statin is more effective in LDL-cholesterol lowering, triglyceride lowering, usCRP lowering, and HDL-C raising than a statin alone.

The NCEP has always suggested beginning LDL-cholesterol lowering with a statin; this should be continued. If a clinician has attained an LDL cholesterol goal with the use of ezetimibe alone without ever having tried a statin and becomes concerned about the results of the ENHANCE study, the clinician could change to statin therapy or the more expensive, bile-acid binding resins.

For the time being, the ENHANCE study may push us more in the direction of increasing statin dosage than in adding ezetimibe to a statin to reach difficult NCEP LDL cholesterol goals. But in many cases, LDL cholesterol goals cannot be reached with statins alone, and with higher dose statins, patients occasionally experience intolerable side effects. In such cases, this one study in this unusual population should not undermine our attempts to reach NCEP LDL-cholesterol goals. Our policy for the moment, and until more results come in from clinical trials, is to attain NCEP LDL-cholesterol goals by any means possible including the use of ezetimibe and simvastatin/exetimibe where necessary.

REFERENCES

1. Kastelein JJ, Sager PT, de GE, Veltri E. Comparison of ezetimibe plus simvastatin versus simvastatin monotherapy on atherosclerosis progression in familial hypercholesterolemia. Design and rationale of the Ezetimibe and Simvastatin in Hypercholesterolemia Enhances Atherosclerosis Regression (ENHANCE) trial. *Am Heart J* 2005; 149: 234–239.
2. Smilde TJ, van WS, Wollersheim H, et al. Effect of aggressive versus conventional lipid lowering on atherosclerosis progression in familial hypercholesterolaemia (ASAP): a prospective, randomised, double-blind trial. *Lancet* 2001; 357: 577–581.
3. Berenson A. Study reveals doubt on drug for cholesterol. *New York Times* Jan 15 2008; A:1.

4. Heartwire. www.theheart.org. Accessed February 16, 2008, internet communication.
5. Taylor AJ, Sullenberger LE, Lee HJ, et al. Arterial Biology for the Investigation of the Treatment Effects of Reducing Cholesterol (ARBITER) 2: a double-blind, placebo-controlled study of extended-release niacin on atherosclerosis progression in secondary prevention patients treated with statins. *Circulation* 2004; 110: 3512–3517.
6. Pearson T, Ballantyne C, Sisk C, et al. Comparison of effects of ezetimibe/simvastatin versus simvastatin versus atorvastatin in reducing C-reactive protein and low-density lipoprotein cholesterol levels. *Am J Cardiol* 2007; 99: 1706–1713.
7. Gagne C, Bays HE, Weiss SR, et al. Efficacy and safety of ezetimibe added to ongoing statin therapy for treatment of patients with primary hyper cholesterolemia. *Am J Cardiol* 2002; 90: 1084–1091.
8. Bays HE, Ose L, Fraser N, et al. A multicenter, randomized, double-blind, placebo-controlled, factorial design study to evaluate the lipid-altering efficacy and safety profile of the ezetimibe/simvastatin tablet compared with ezetimibe and simvastatin monotherapy in patients with primary hypercholesterolemia. *Clin Ther* 2004; 26: 1758–1773.
9. Aminbakhsh A, Mancini GB. Carotid intima-media thickness measurements: what defines an abnormality? A systematic review. *Clin Invest Med* 1999; 22: 149–157.
10. van WS, Smilde TJ, Trip MD, et al. Long-term safety and efficacy of high-dose atorvastatin treatment in patients with familial hypercholesterolemia. *Am J Cardiol* 2005; 95: 264–266.
11. Salonen R, Nyyssonen K, Porkkala E, et al. Kuopio Atherosclerosis Prevention Study (KAPS). A population-based primary preventive trial of the effect of LDL lowering on atherosclerotic progression in carotid and femoral arteries. *Circulation* 1995; 92: 1758–1764.
12. Kane JP, Malloy MJ, Ports TA, et al. Regression of coronary atherosclerosis during treatment of familial hypercholesterolemia with combined drug regimens. *JAMA* 1990; 264: 3007–3012.
13. Dalager S, Paaske WP, Kristensen IB, et al. Artery-related differences in atherosclerosis expression: implications for atherogenesis and dynamics in intima-media thickness. *Stroke* 2007; 38: 2698–2705.
14. Wang DQ. Regulation of intestinal cholesterol absorption. *Annu Rev Physiol* 2007; 69: 221–248.
15. Santosa S, Varady KA, AbuMweis S, Jones PJ. Physiological and therapeutic factors affecting cholesterol metabolism: does a reciprocal relationship between cholesterol absorption and synthesis really exist? *Life Sci* 2007; 80: 505–514.

EVIDENCE-BASED MEDICINE

Deciding Whether To Screen Patients for Abdominal Aortic Aneurysm

Thomas McGinn MD, MPH and Deborah Korenstein, MD

Division of General Internal Medicine, Department of Medicine, Mount Sinai School of Medicine,
New York, NY

This is the first of a series of articles on evidence-based medicine in the *Mount Sinai Journal of Medicine*, emphasizing applications of concepts and procedures of evidence-based medicine to clinical practice. The articles will focus on the retrieval, interpretation, and application of evidence related to a current clinical problem. The first discussion is based on the following case.

THE CASE

A 67-year-old man makes his annual visit to a physician in general medicine practice. The patient says that he feels well, but he is worried about the possibility that he may have an abdominal aortic aneurysm (AAA). Recently, his close friend and partner of over 30 years died suddenly from an AAA rupture. Like the patient, his friend was in his late 60s and had a history of mild hypertension and hyperlipidemia. "If my friend could get an aneurysm, so could I," the patient says. The physician, aware that the patient's hypertension and hyperlipidemia have been adequately controlled with a diuretic and a statin, that he has no other medical or family history of significance, and that he has a decent diet, exercises on a regular basis, and has never smoked, tells the patient that his risk of an AAA is very small. "I want to be sure," the patient insists. The physician is not certain that the patient's baseline risk warrants screening him for AAA; he is also concerned that, if the test produces a false positive result or shows slight changes warranting no immediate intervention, it may cause the patient's anxiety to increase rather than abate. "Let me think it over," the physician says. "I'll get back to you."

Address Correspondence to:

Thomas McGinn MD, MPH
and Deborah Korenstein, MD
Division of General Internal
Medicine
Department of Medicine
Mount Sinai School of Medicine
New York, NY
Email: thomas.mcginn@
mountsinai.org

FOCUSED CLINICAL QUESTION

To determine whether an AAA screening would benefit this patient, the physician decides to review the relevant medical literature. He begins by formulating a focused question consisting of 4 elements: the patient (P), the intervention (I), the comparison (C), and the outcome (O). This is often called PICO.[1] The explicit inclusion of these 4 components allows for a more efficient and directed search of the literature and helps to clarify the exact problem of interest. In this case, the question's components are as follows: in asymptomatic adults over 65 (P), does screening ultrasound for AAA (I), in comparison with the usual care (C), reduce mortality and improve quality of life (O)?

SEARCH OF THE LITERATURE

Performing a literature search in PubMed, the physician enters the search term "aneurysm and screening." This search brings up approximately 45,000 citations. He repeats the search using the "clinical queries" field under the "treatment" category, which focuses the search on higher quality

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20035

© 2008 Mount Sinai School of Medicine

primary evidence by the addition of "filtering" terms such as "randomized controlled trial" (RCT) for treatment articles.[1] This strategy results in 250 citations, which are still too many to review. To narrow the search further, the physician uses the "limits" fields to restrict the search to English language, human subjects, and core journals. This brings the number of citations down to a manageable list of 65 citations, which the physician decides to review.

Among the citations are 2 large RCTs of ultrasound screening for AAA. One is a landmark 2002 study, published in the *Lancet*, involving approximately 67,000 men ages 65 to 74; the second is a 2005 trial, published in the *British Medical Journal*, involving 12,000 men who were followed for an average of 52 months.[3] Both trials compared the offer of a screening procedure with the absence of an offer.[3]

RCTs are becoming the more common method for evaluating screening and diagnostic tests. Customarily, such tests are evaluated through a comparison of their results with a reference standard. This approach shows a test's operating characteristics, such as its accuracy, positive and negative predictive values, and likelihood ratios, but it cannot show the impact of its use on mortality or quality of life. The actual clinical impact of a screening test can be demonstrated only with an RCT.

CRITICAL APPRAISAL OF THE LITERATURE

A critical appraisal of a clinical trial has 3 main components: an assessment of the trial's validity, an analysis of the results, and a review of the study's applicability. For the sake of discussion, we will focus on the physician's critical appraisal of the 2005 RCT cited previously.[4]

Validity Criteria for Screening Trials

Appraising the 2005 trial, the physician finds that it met most of the basic validity criteria for an intervention trial:

- The study was randomized and controlled, the unit of randomization being the "offer" of screening sonography versus no offer. However, the physician notes that, although the randomization appears to have been properly conducted, the study includes no description of the frequency of risk factors and comorbidities in the 2 groups. The equal distribution of patient characteristics would have ensured that the randomization was balanced and helped to identify important risk factors for AAA within the study population, which may ultimately help in identifying appropriate patients for screening.
- The study was performed and analyzed in an intention-to-treat fashion. For example, subjects in the sonography arm who were offered screening but did not actually undergo sonography were still counted in the screened group.
- There appears to have been appropriate blinding of those who assessed the primary endpoint, death secondary to ruptured AAA. This is an extremely important aspect of the study because assessing whether death was caused by rupture from an AAA or from other causes entails judgment and some aspect of subjectivity. In this case, only 6% of patients had autopsies.

Results

Consulting a review of both the 2002 and 2005 RCTs by the ACP Journal Club (a filtered database), the physician finds that the AAA screening test consistently reduced a ruptured AAA as a cause of death.[5,6] Although there was no difference between the screened and unscreened groups with respect to all-cause mortality, there was a 67% relative risk reduction (RRR) of mortality caused specifically by AAA; this result was statistically significant. At the same time, the event rates were small; in the screened group, the rate was 0.14%, and in the nonscreened group, it was 0.43%. The studies also showed that, given the event rates, the number of patients who would have to be screened in order to prevent 1 death from AAA over a period of 52 months [the number needed to screen (NNS)] is 350. This means that 1 death from AAA is prevented for every 350 patients (similar to the study patients) who are screened. NNS corresponds to the number needed to treat (NNT) in a traditional treatment study.

Applicability

The physician determines that the 2005 study applies to his patient because the patient would have met the entry criteria.

THE PHYSICIAN'S DECISION

The physician notes that, in the studies under review, screening for AAA was associated with a statistically significant benefit–a 67% RRR in mortality caused by AAA–but the event rates in the screened and unscreened groups were small, and this raises the question of whether the difference was also significant clinically.

This question is not uncommon. Table 1 illustrates the interplay between the RRR associated with a screening test and its clinical benefit. The first column shows baseline event rates without screening. The RRR associated with a screening strategy remains constant, and the remainder of the table demonstrates the clinical impact of screening the different risk groups. As the table illustrates, as the baseline risk declines, the absolute number of events prevented (absolute risk reduction) also declines, and the NNS rises, reflecting the smaller absolute benefit.[7] In a lower risk group, a large RRR can be associated with a very small absolute risk reduction; looking only at the RRR may lead to an overestimation of the clinical benefit.

The physician also notes that, compared with NNTs in most treatment studies, the NNS in the 2005 study of the AAA screening test (350) was fairly high. However, a high NNS is more common and acceptable in evaluations of screening tests than in traditional treatment studies because the patients have a relatively low risk of death. For example, the NNS for annual mammography in women over the age of 50 to prevent death from breast cancer is approximately 1000,[8] and the NNS with fecal occult blood testing to prevent death from colon cancer is approximately 500.[9]

In general, the determination of whether an NNS is acceptable is a complex process, involving the integration of several factors such as the cost and invasiveness of the screening test, the false positive rate and implications for follow-up tests, and the severity and treatability of the illness at hand.

Despite some marginal data, the physician concludes that an AAA screening test can provide his patient with a significant benefit. In light of this and the patient's anxiety concerning a possible AAA, the physician decides to recommend the test. The patient readily agrees, and the test is performed. The results are negative for aneurysm, relieving both the patient and the physician.

Table 1. *Interplay Between Event Rates and RRR.*

Event Rate: No Screen (%)	RRR (%)	Event Rate: Screened (%)	ARR (%)	NNS
60	33	40	20	5
6	33	4	2	50
0.6	33	0.4	0.2	500

Abbreviations: ARR, absolute risk reduction; NNS, number needed to screen; RRR, relative risk reduction.

CONCLUSIONS

This case illustrates several important points:

- Compared with the traditional method for evaluating a new screening test, which involves comparing the test results to a reference standard, an RCT is far superior, providing insight into the test's true benefits. An RCT is, however, far more expensive and time-consuming to perform.
- Screening and prevention trials typically require large numbers of patients because of the low rate of the outcome event, which in this case is a ruptured AAA.
- The combination of a large sample size and a low number of outcome events may result in statistically significant results that are not significant clinically.
- NNT(s) in screening studies tends to be higher than standard NNTs from treatment trials. An NNS of 350 for AAA is within acceptable range.

In future articles on evidence-based medicine, we will discuss other aspects of the evidence cycle. The complete cycle is shown here:

1. A case and dilemma.
2. A well-formulated clinical question.
3. A search of the literature to answer the question.
4. A critical appraisal of the literature.
5. The incorporation of patient values and preferences.
6. Finally, the application of the results to the care of the patient.

Each article in the series will focus on 1 or 2 specific areas, such as methods for searching the medical literature for meaningful data and the procedure for concealment in medical research. Although not meant to be exhaustive, the series will demonstrate how a physician can obtain reliable evidence-based answers to specific clinical questions of importance to his or her patients.

REFERENCES

1. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 1995; 123: A12–A13.
2. Ashton HA, Buxton MJ, Day NE, et al., and the Multicentre Aneurysm Screening Study Group. The Multicentre Aneurysm Screening Study (MASS) into the effect of abdominal aortic aneurysm screening on mortality in men: a randomised controlled trial. *Lancet* 2002; 360: 1531–1539.
3. Lindholt JS, Juul S, Fasting H, Henneberg EW. Screening for abdominal aortic aneurysms: single centre randomised controlled trial [published correction appears in *BMJ* 2005; 331: 876]. *BMJ* 2005; 330: 750.

4. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1993; 270: 2598–2601.
5. Cina CS. Screening for abdominal aortic aneurysm reduced death from AAA in older men. *ACP J Club* 2003; 138: 66.
6. Michaels JA. Screening for abdominal aortic aneurysm (AAA) reduced mortality caused by AAA in Danish men 64 to 73 years of age. *ACP J Club* 2005; 143: 39.
7. Barratt A, Wyer PC, Hatala R, et al., and the Evidence-Based Medicine Teaching Tips Working Group. Tips for learners of evidence-based medicine: 1. Relative risk reduction, absolute risk reduction and number needed to treat. *CMAJ* 2004; 171: 353–358.
8. Humphrey LL, Helfand M, Chan BK, Woolf SH. Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med* 2002; 137(pt 1): 347–360.
9. U.S. Preventive Services Task Force. USPTF screening for colorectal cancer: recommendation and rationale. *Ann Intern Med* 2002; 137: 129–131.

DOI:10.1002/MSJ

ETHICS

Capacity Assessment and Intervention in 56-yr-Old Man with Schizophrenia

Alan Schlechter, MD

Fellow, NYU Child Study Center, Visiting Associate Faculty, Mount Sinai School of Medicine, New York, NY

ABSTRACT

The importance of the doctor-patient relationship among patients with schizophrenia is especially evident within the context of medication adherence. The association between patients' positive perception of their psychiatrist and medication adherence has been established in several studies. This is the case of a 56-year-old man with schizophrenia and multiple medical problems who is refusing to take any medications. We discuss the risks and benefits of having a patient declared incompetent along with assessing the patient's capacity. The case highlights situations when a physician might question a patient's decisional capacity but conclude that the decision should be accepted so as to maintain rapport with the patient and continue treatment. ***Mt Sinai J Med 75:153–155, 2008.*** © *2008 Mount Sinai School of Medicine*

Key Words: schizophrenia, medication adherence, medical ethisc, decisional capacity.

When assessing a patient's delusional capacity, a physician makes a clinical judgment that may impact the patient's autonomy, confidentiality, and relationship with other physicians. To have capacity, that is, to be autonomous, a patient must demonstrate an understanding of the proposed treatment, an appreciation of the risks and benefits involved, and then make a decision consistent with the patient's values and circumstances. A physician might need to question relatives or close friends of the patient to ascertain the patient's baseline mental status, and sometimes does so without the patient's consent. If a physician takes the matter before a judge and the patient is declared incompetent, the patient may lose trust in physicians and future work with the patient might be compromised. There will be situations when a physician might question a patient's decisional capacity but may conclude that the decision should be accepted so as to maintain rapport with the patient and continue treatment. Physicians assessing capacity must weigh the risks and benefits of having a patient declared incompetent along with assessing the patient's capacity.

Address Correspondence to:

Alan Schlechter
Fellow, NYU Child Study Center
Visiting Associate Faculty
Mount Sinai School of Medicine
New York, NY
Email: alan.schlechter@mssm.edu

CASE REPORT

Mr Sherman was a 56-yr-old man with a past medical history of chronic obstructive pulmonary disease (COPD) with multiple exacerbations requiring hospitalization. He had hypertension, type 2 diabetes, hepatitis C, chronic renal insufficiency, and a past cerebral vascular accident. He also had a diagnosis of schizophrenia. On this occasion, Mr Sherman was admitted because of a COPD exacerbation and his sugar level was found to be greater than 300 mg/dl. Throughout his admission, he denied having any medical or psychiatric illness. He was adhering to all medications while in the hospital but said he would not need follow-up or medications as an outpatient. I was on the psychiatry consultation and liaison service and was called to see whether Mr Sherman had the capacity to understand his medical condition and whether he was safe to be discharged.

On entering the room he was well groomed, cooperative, calm, and easy to talk with. He was oriented to person, place, time, and aware of national events. He did not appear guarded or

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20029

© 2008 Mount Sinai School of Medicine

internally preoccupied. He had a full affect and said his mood was good except for his difficulty in breathing, which he related to his time in the military as he was formerly a "frog man". Mr Sherman categorically denied having any illness, either medical or psychiatric. He stated that he had had a test in the past, and it showed he did not have diabetes. He said there was no way anyone could prove he had diabetes because of this previous test.

Mr Sherman denied suicidal or homicidal ideation and auditory or visual hallucinations. He denied having any paranoid symptoms and did not seem guarded. He was told he could suffer serious harm if his diabetes was not treated, including kidney failure, loss of limbs, stroke, and death. He responded, "that will happen to all of us one day".

His past psychiatric history according to his medical chart included his first psychiatric hospitalizations during 1977–1979 at the Bronx VA for psychotic episodes with auditory hallucinations, visual hallucinations, paranoid ideation, mutism, violent behavior, and agitation. He was treated with thorazine, haldol, and most recently risperidone, all of which he decided to stop taking. With Mr Sherman's permission we contacted his son and wife who stated that no antipsychotic drug had ever been helpful. From 1979 until 1996, the patient did not have psychiatric follow-up or hospitalizations.

Mr Sherman's wife was legally separated from him in 1988, but she lives a short distance away and visits him every day with their son. Mr Sherman has a phone with which he can call his wife or son if necessary. She is unaware of any suicidal, homicidal, or self-injurious ideation on the part of her husband. Both Mr Sherman and his family confirm that he lives a relatively happy life, spending his days with his friends and family.

DISCUSSION

Mr Sherman did not acknowledge any illness besides a "breathing difficulty". He did not show an ability to weigh the risks and benefits of not taking his medications. He was able to say that he did not want to take any medication, but he justified his choice on false premises. To sum up, he did not display the capacity for making a medical decision about his future medications.

If Mr Sherman's death had been imminent, medications would have been imposed on him. Owing to the chronic nature of his illness, nothing untoward might happen in the next day, month, or year. The inability to predict when a dire consequence will result diminishes the need for forcing treatment on this patient. Patients often make poor health decisions when they deny the potential harm to themselves. Mr Sherman's situation is more complex, given his global lack of awareness about his health status. A meaningful comparison would be to that of allowing someone to drive while intoxicated. An intoxicated driver does not appropriately weigh the risks and benefits of driving while intoxicated or proceeds under the false premise that he/she is not intoxicated and perfectly capable of driving safely.

Mr Sherman's diagnosis of schizophrenia gives rise to the questions of whether he was delusional, in denial, or cognitively impaired. A person who is under a delusion cannot be persuaded with the facts, while someone in denial might be amenable to realize what he/she is unaware of. During the course of treating Mr Sherman, I began using the word "sugar" instead of "diabetes" to discuss his illness. He was more amenable to accepting this term and willing to discuss his "sugar" problem though he continued to refuse taking any medications on discharge. By altering the language of illness I was able to make some headway with Mr Sherman. To truly assess whether a patient is delusional or in denial takes time and involves the creation of rapport.

One approach to Mr Sherman may have been to try a course of antipsychotic to see whether his "delusion" would improve. Antipsychotics nominally treat the cognitive deficits associated with schizophrenia if at all, and have a very limited efficacy with long-term delusions. I knew that he had been on antipsychotics in the past without success. Furthermore, if Mr Sherman was forced to take antipsychotics the imposition could be expected to have a negative effect on the doctor–patient relationship which would undermine future adherence to treatment plans. Forcing Mr Sherman to take his medications would have meant placing him in some kind of supported housing situation and removing him from his home and community. This radical intervention would have been medically safer, but it would have a dramatic impact on his quality of life.

The original consult was to assess Mr Sherman's capacity, but the underlying need was for creating a rapport with him and with his medical team. Over the course of a week, I spent 30 minutes daily discussing Mr Sherman's feelings about his health. I was supportive and encouraged him to follow-up with a psychiatrist who could continue to monitor his mental and physical health. The psychiatrist will hopefully develop a trusting relationship with Mr Sherman and seize opportunities to provide him with needed medical care. I explained that our conversations had been friendly and made it clear that medications would not be imposed upon him. Mr Sherman was amenable to his outpatient visits with the psychiatrist and was on time for his first follow-up appointment 1 week after discharge.

When this case was presented at an ethics discussion, there were many questions concerning the acceptability of this outcome. Mr Sherman has type 2 diabetes and is in danger of losing a limb, worsening renal failure, stroke, and death. He did not have the capacity to understand either his psychiatric or medical disease or his need for treatment. Still, it would be a dramatic and paternalistic intervention to remove Mr Sherman from the life that he enjoys in order to assure that he takes medications which may or may not improve his mental functioning or prolong his life. Our hope was to monitor Mr Sherman more closely and to help him accept some part of his illness or treatment over time. In the case of chronic mental illness, similar to a chronic medical condition, we must set realistic goals and balance the risks and benefits of the patient's condition and quality of life. In this case, Mr Sherman returned to the hospital as an outpatient. He has retained some measure of his autonomy and there is a sign that the doctor–patient relationship has been strengthened. We continue our work from here.

NEW DRUGS AND THE FDA

Neurotherapeutics in Multiple Sclerosis: Novel Agents and Emerging Treatment Strategies

Tracy DeAngelis, MD and Fred Lublin MD

Department of Neurology, Mount Sinai School of Medicine, New York, NY, USA

ABSTRACT

New insights into the complex immunopathogenesis of multiple sclerosis have led to a proliferation of promising new therapeutic strategies. While the current armamentarium of immunomodulatory medications has demonstrated beneficial effects on the disease, more effective and tolerable therapies are needed. Several novel therapeutic strategies under investigation include oral therapies, monoclonal antibodies, symptomatic treatments, insights into neuroprotection and repair as well as combination regimens. New therapies may prove more efficacious and tolerable than the available arsenal of treatments; however, decisions regarding first-line therapies will expectedly become more complicated, with greater influence if risk-to-benefit ratios in light of premature safety data. Biomarker profiles may help elucidate disease subtypes as well as therapeutic response in an effort to individualize treatment choice. This review will highlight recent promising therapeutic strategies under investigation in the field of MS. ***Mt Sinai J Med 75:157–167, 2008.*** *© 2008 Mount Sinai School of Medicine*

Address Correspondence to:

Tracy DeAngelis and Fred Lublin
Corinne Goldsmith Dickinson Center for Multiple Sclerosis
New York, NY, USA
Email: tracy.deangelis@mssm.edu;
fred.lublin@mssm.edu

Key Words: multiple sclerosis, interferon beta, glatiramer acetate, fingolimod, fumarate, cladribine, teriflunomide, natalizumab, alemtuzumab, daclizumab, rituximab, fampridine, antigen-based therapies, statins.

The past decade has marked the advent of various new therapeutic strategies in multiple sclerosis (MS). The vigorous growth seen in the field of MS research serves as a neurotherapeutic paradigm for several other chronic neurologic diseases characterized by inflammatory and neurodegenerative processes, for example, Alzheimer disease and Parkinson disease. Such diseases face the therapeutic challenge of addressing complex pathophysiology and unpredictably variable clinical spectra. In addition, they share the struggle to preserve neurologic function before disease progression outruns therapeutic effect. In this scenario, earlier diagnosis and initiation of therapy likely plays a critical role in improving patient outcome. Current advances in MS neurotherapeutics are attempting to address these complicated issues and move promising benchwork to the bedside.

MS is a chronic disease of the central nervous system (CNS), which historically has been considered an autoimmune inflammatory disorder. To date, the etiology of MS remains unknown; however, it likely results from a breakdown of immune tolerance related to genetic predisposition, environmental triggers, and immune dysregulation. Our broadening understanding of the immunopathogenesis of MS continues to foster clinical research. Pathologic studies support the presence of both inflammation and neurodegeneration, even in acute plaques, which contain activated macrophages and T cells as well as B cells, demyelinated axons, and axonal injury.[1] Transition from the inflammatory to the progressive phase may result from immune

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20030

© 2008 Mount Sinai School of Medicine

dysregulation secondary to inflammatory activity or even potentially subsequent to a loss of certain inflammatory mechanisms. Recent research has shed new insight into this complex pathophysiology, including the role of specific immune cells and cytokines, as well as insights into remyelination and repair.

Despite an armamentarium of immunomodulatory medications that have demonstrated beneficial effects on disease activity, more effective and tolerable therapies are needed for MS patients. A multitude of novel therapeutic strategies currently being tested include oral therapies, monoclonal antibodies (mAbs), unique symptomatic treatments, and there have been new insights into neuroprotection and repair. In addition, methods to enhance existing therapies are under investigation, for example, early initiation of treatment, as well as various combination regimens. New therapies, along with variations of available treatments, may prove more efficacious and tolerable than the available arsenal of treatments. Nevertheless, as the treatment horizon broadens, choosing first-line therapies will become more complicated, with greater influence on risk-to-benefit ratios in light of premature safety data. Disease heterogeneity and our limited ability to prognosticate will undoubtedly pose obstacles in the application of novel agents. Importantly, biomarker fingerprints, both clinical and paraclinical, have potential to elucidate disease subtypes as well as response to therapy in an effort to individualize treatments. A complete discussion of all research currently under way is beyond the scope of this review, will highlight recent promising neurotherapeutics in the field of MS.

DISEASE-MODIFYING THERAPY

Enhancing Existing Therapies

After more than a decade, the introduction of disease-modifying therapies (DMTs) has clearly made a significant impact on the treatment of MS. At present, 6 available DMT agents have been approved by the Food and Drug Administration (FDA) for the treatment of relapsing-remitting MS (RRMS). These include interferon (IFN)-beta preparations, glatiramer acetate (GA), the mAb, natalizumab, and the chemotherapeutic agent, mitoxantrone (MTX). Although there is clear, convincing evidence demonstrating their efficacy in decreasing clinical and MRI disease activity, the DMTs are, nonetheless, partly effective, and their long-term impact on disease progression remains unclear. Moreover, their parenteral administration and side effects often impede patient adherence.

On average, all injectable DMTs have demonstrated a beneficial effect on decreasing annual relapse rates by approximately 30% when compared with placebo. Nevertheless, the caution to compare efficacy across clinical trials provides limited ascertainment of therapeutic superiority. DMTs have very distinct mechanisms of action. IFNs modulate antigen presentation via the down-regulation of costimulatory molecules. Although they do not exert direct effects within the CNS, their action in the periphery limits T-lymphocyte trafficking into the CNS by effects on matrix metalloproteinases and adhesion molecules. Furthermore, they reduce permeability of the blood-brain barrier (BBB). GA's mechanism of action involves production of GA-reactive T2 cells that cross the BBB and exert direct effects on the CNS inflammatory milieu, specifically increasing secretion of suppressor-type cytokines and down-regulating inflammation.

Several studies have aimed to build upon the current DMTs to enhance efficacy, tolerability, and adherence. For example, initiating DMTs earlier appears to have a significant benefit on patient outcome. Additionally, recent head-to-head data have suggested that the efficacy of IFNs and GA may be similar, enabling clinicians to enhance adherence by tailoring treatments to individual patient preferences in terms of side effects and dose regimens. Combination therapies as well as alternative dosing regimens may also enhance efficacy of current treatments.

Early Initiation of DMT

To date, several studies have accumulated evidence that supports early initiation of DMT. Pathologic studies have demonstrated evidence of axonal injury occurring in acute MS plaques,[2] and inflammatory axonal damage may occur even in the earliest stages of the disease.[3] The extension studies of the pivotal GA trial and PRISMS trials both demonstrate that a 2-year delay in treatment can make a critically significant impact on development of disease progression.[4–6] Likewise, examining populations presenting with a first episode of MS, referred to as clinically isolated syndromes (CIS), has demonstrated the benefit of early therapy on risk of conversion to clinically definite MS (CDMS). For example, results of the Early Treatment of MS Study (ETOMS), Controlled High-Risk Subjects Avonex Multiple Sclerosis Prevention Study (CHAMPS), and most recently, the Betaseron in Newly Emerging Multiple Sclerosis for Initial Treatment (BENEFIT) trials clearly demonstrate that early initiation of IFN-beta is effective in delaying conversion to CDMS.[7–10]

Although there is currently no available data on GA's effect on CIS, a randomized, placebo-controlled trial is currently in progress.[11]

Arguments against early initiation of DMT have generally focused on the inability to predict disease course and the existence of a "benign" form of MS, which potentially may not require treatment with DMTs. Importantly, a recent longitudinal study of patients believed to have a "benign" form of MS at 10 years revealed that 2 decades later, 21% had progressed to require the use of a cane for ambulatory assistance.[12] As such, not treating this population based on their "benign" label may have negative consequences, and better criteria are needed to identify which patients will remain with mild disability and which will progress.

Head-to-Head Comparisons of DMTs

Until recently, no blinded, randomized, controlled trials of 2 years or longer had compared any of the DMTs. Three recent studies, Rebif versus Glatiramer Acetate in Relapsing MS Disease (REGARD), Betaseron versus Copaxone in MS with Triple-Dose Gadolinium and 3-T MRI Endpoints (BECOME) and Betaseron Efficacy Yielding Outcomes of a New Dose (BEYOND), have provided head-to-head data on the comparative efficacy of IFNs versus GA.

The REGARD trial was a 96-week study of 764 patients with relapsing MS randomized to receive either IFN-beta 1a 44 mcg injections thrice weekly versus GA.[13] The primary outcome measure was time to first relapse, and results demonstrated no difference between groups. Additionally, 460 of the 764 patients underwent MRI every 6 months for secondary and tertiary outcome measures. No difference in the number of new or enlarging T2 lesions per patient per scan was noted between the GA-treated (0.7 lesions) and IFN-treated (0.6 lesions) groups. A statistically significant reduction of 50% in the number of gadolinium (Gd)-enhancing lesions per patient per scan, along with a significant reduction in combined unique lesions (Gd-enhancing lesions and new nonenhancing T2 lesions), were found in the IFN-treated group. Such a difference appears consistent with IFN's reduction of BBB permeability. With respect to the tertiary outcome measure, annualized relapse rate, no difference was identified between groups, with 0.30 relapses per patient per year in the IFN-treated group, and 0.29 relapses per patient per year in the GA-treated group.

Of interest, the annualized relapse rate in this study, as well as several recent trials,[14] was significantly lower in comparison to the pivotal trials for both IFN-beta and GA. Specifically, there were 45% fewer relapses than anticipated by investigators. The significance of this low relapse rate as well as that observed in other trials remains unclear; however, it may reflect a changing disease population, likely influenced by early diagnosis and treatment. Importantly, the low event rate in REGARD may have rendered it statistically underpowered to detect differences between treatment groups. With this in mind, future trials may require larger populations to detect significant differences between treatments, imposing temporal and economic constraints on the design of MS research studies.

Another trial that involved a head-to-head comparison between IFN and GA was BECOME, the first investigator-initiated, single-site, randomized, prospective, rater-blinded trial that directly compared these agents. BECOME was a 24-month-study of 75 patients with relapsing MS or CIS randomized to IFN-beta 1b or GA.[15] Patients underwent monthly triple-dose Gd-enhanced 3-Tesla MRI for 24 months. Unlike REGARD, which identified a significant difference in MRI outcome measures in favor of the IFN-treated group, BECOME surprisingly found no difference in the primary outcome measure of average number of combined active lesions (Gd-enhancing lesions plus new nonenhancing T2 lesions) per patient per scan between treatments. Whether or not this relates to the MRI technique used in this study compared with the standard 1.5-T MR with single-dose Gd remains unclear, but it certainly raises issues about the sensitivity of conventional imaging to characterize disease activity radiographically. Additionally, there were no differences between the 2 groups in any of the secondary clinical outcomes, annualized relapse rate and disability progression.[16] Of note, however, the study was not powered to detect significant clinical changes.

The BEYOND study was designed to investigate the efficacy, tolerability, and safety of double-dose (500 mcg) IFN-beta 1b versus the standard 250 mcg dose versus a comparative GA-treatment arm. The primary outcome measure, time to first relapse, was similar in all 3 treatment arms with no statistically significant superiority of the 500 mcg IFN-beta 1b dose over the 250 mcg dose or GA.[17] Again, in contrast to previous trials, a very low relapse rate was noted in this study population.

Although these head-to-head data are extremely helpful in rationalizing the use of all DMTs, individual response to a particular agent remains unknown. As such, choice of therapy ultimately rests on a thorough review of different dose regimens and side effect profiles, so patients can opt for a DMT with which they will comply and tolerate.

Combination Therapies and Alternative Dosing Regimens

Combinations of conventional DMTs and/or additions of immunosuppressive agents to DMT may provide synergistic effects to treatment regimens that could potentially enhance patient outcome. Agents considered for combination therapy generally have unique mechanisms of action with the potential to provide cumulative or synergistic efficacy. As such, mechanistic differences between the IFN and GA position these agents as an attractive cocktail. A 6-month pilot trial of 32 patients with RRMS assessed the safety of IFN-beta 1a and GA and compared the number of Gd-enhancing lesions in patients who were receiving combination versus either treatment alone.[18,19] No adverse events (AEs) or any increases in Gd-enhancing lesions were noted. Currently, a 3-year NIH-funded, randomized, blinded trial, Combi-Rx, is comparing relapse-rate reduction between 3 treatment arms, IFN-beta 1a 30 μg intramuscularly (IM) weekly, GA 20 mg subcutaneously daily, versus a combination of IFN and GA. Results are expected in 2011.

In addition to combinations of DMTs, combination-induction strategies that use immunosuppressants and DMTs are also of interest, for example, MTX-induction therapy before initiation of DMT with IFN or GA. MTX, an FDA-approved treatment for relapsing forms of SPMS and worsening RRMS, is an anthracenedione DNA intercalator whose immunomodulatory mechanisms include an antiproliferative effect on T cells as well as suppression of humoral immunity. The rationale supporting this approach involves an initial priming of the immune system to enhance the action of the maintenance DMT.

Importantly, whereas immunomodulatory and/or immunosuppressant cocktails offer a potentially attractive therapeutic approach, the possible risk of unexpected adverse events, not evident with monotherapy, must be emphasized. A recent example is the appearance of progressive multifocal leukoencephalopathy (PML) in patients who are receiving combination treatment with natalizumab and IFN-beta 1a, which will be discussed in detail in a subsequent section.

ORAL THERAPIES

Several new oral therapies are currently under investigation. Should long-term studies demonstrate similar or better efficacy and safety compared with current DMTs, novel oral therapies will expectedly pose attractive first-line agents for MS patients reluctant to comply with injectable treatments. Improvement in long-term adherence offers the important potential to optimize therapeutic outcomes.

Fingolimod

FTY720, or fingolimod, is a promising new oral immunomodulatory agent under investigation for relapsing MS. It is a synthetic analog of sphingosine 1 phosphate (S1P), which modulates S1P receptors on thymocytes and lymphocytes to sequester lymphocytes in secondary lymphoid organs, inhibiting their egress to the CNS. In a recent Phase II trial lead by Kappos *et al*,[20] fingolimod given once daily demonstrated a significant and rapid reduction in MRI measures of inflammation and in relapse-related clinical endpoints in MS patients. In this double-blind, randomized, placebo-controlled study, 281 patients with relapsing MS were randomized to receive fingolimod (1.25 mg or 5 mg orally/daily) or placebo with a primary endpoint measure of total Gd-enhancing lesions on MRI at monthly intervals for 6 months. The mean total number of Gd-enhancing lesions was lower in both treatment groups (1.25 mg: 1 lesion, $P < .001$; 5 mg: 3 lesions, $P < .006$) compared with placebo (5 lesions). In addition, the annualized relapse rate was 0.77 in the placebo group compared with 0.35 in the 1.25-mg fingolimod group ($P = .009$) and 0.036 for the 5-mg fingolimod group ($P = .01$). In the extension phase, both the number of Gd-enhancing lesions as well as relapse rates remained low in both fingolimod-treatment groups, and both measures decreased in patients switched from placebo to fingolimod.

Adverse events associated with fingolimod included nasopharyngitis, dyspnea, headache, diarrhea, and nausea. A case of posterior reversible leukoencephalopathy syndrome (PRES) occurred in the 5-mg treatment group. In addition, lymphopenia, clinically asymptomatic elevations in liver alanine aminotransferase, bradycardia, and pulmonary function test abnormalities were also reported. Although results of this study are encouraging, 2 larger Phase III trials are currently under way, including 1 comparing fingolimod to an approved DMT. Of interest, *in vitro* cell culture studies have shown fingolimod to increase both mature and immature oligodendrocyte populations, suggesting a unique potential impact on remyelination.[21,22]

Fumaric Acid

Fumaric acid is an immunomodulator used for the treatment of psoriasis. Novel fumaric acid esters with better tolerability have recently been developed. An oral fumaric acid derivation, BG00012, is currently being studied in 2 Phase III trials in patients with RRMS.[23] *In vitro* studies demonstrate that fumaric acid inhibits expression of proinflammatory adhesion molecules and cytokines as well as exhibiting neuroprotective effects.[24,25] A 24-week Phase II double-blind placebo-controlled trial comparing BG00012 (120 mg, 360 mg or 720 mg) daily to placebo demonstrated a significant reduction of Gd-enhancing lesions in the highest-dose BG00012 treatment group.[26] Moreover, the BG00012 group demonstrated 48% fewer T2 lesions and 53% fewer T1 black holes compared with placebo ($P = .001$). The most common side effects included headache, nasopharyngitis, nausea, and flushing, with no significant difference between the treatment and placebo arms. A large Phase III study on fumaric acid in RRMS is now recruiting.

Cladribine

Cladribine is an adenosine deaminase-resistant purine nucleoside analog that preferentially depletes subpopulations of lymphocytes. Currently, its indications are oncologic, as an effective treatment for hairy-cell leukemia and lymphoma. Preliminary studies of parenteral cladribine in MS demonstrated efficacy in clinical and MRI parameters with a safety profile of doses up to 2.1 mg/kg similar to placebo. Doses of 0.7–2.1 mg/kg can be achieved with oral preparations, and 2 studies that examined the safety and efficacy of oral cladribine are under way. A 2-year Phase III randomized, placebo-controlled trial of cladribine monotherapy in patients with RRMS and expanded disability status scale (EDSS) scores of 0–5.5 has completed enrollment. The primary outcome measure is relapse rate from baseline to Week 96. Another Phase II study combining IFN-beta 1a 44 mcg thrice weekly with oral cladribine is under way with a primary outcome measure of the number of new Gd-enhancing lesions from baseline to week 96.[27]

Teriflunomide

Teriflunomide is a dihydro-orotate dehydrogenase inhibitor, which has demonstrated immunomodulatory effects including the ability to suppress experimental allergic encephalomyelitis. A recent Phase II study that assessed safety and efficacy of oral teriflunomide in relapsing forms of MS found the agent to be well tolerated and effective in reducing MRI lesions.[28] In this 36-week randomized, double-blind, placebo-controlled trial in relapsing forms of MS, 179 patients were randomized to either placebo, teriflunomide 7 mg/day, or teriflunomide 14 mg/day. MRI brain scans were obtained every 6 weeks with a primary endpoint of combined unique active lesions per scan. Secondary endpoints included MRI disease burden, relapse frequency, and disability accumulation. The median number of combined unique active lesions per scan was significantly lower in the treatment groups: 0.5, 0.2, and 0.3 in the placebo, teriflunomide 7 mg/day ($P < .03$ *vs* placebo), and teriflunomide 14 mg/day ($P < .01$ versus placebo) groups, respectively. Moreover, teriflunomide-treated patients also had significantly fewer T1-enhancing lesions per scan, new or enlarging T2 lesions per scan, and new T2 lesions, with the higher-dose group demonstrating a significant reduction in T2 disease burden. Additionally, the higher-dose treatment group exhibited a trend toward lower annualized relapse rates, as well as significantly fewer patients demonstrating an increase in disability. Treatment was well tolerated with similar numbers of adverse events in all treatment groups. Two Phase III studies that investigated the efficacy and safety of teriflunomide in patients with CIS and RRMS are currently recruiting.

HMG-CoA Reductase Inhibitors (Statins)

Interest in HMG-CoA reductase inhibitors, or statins, as potential therapeutic agents in MS is based on several experimental studies that demonstrated a possible anti-inflammatory, potentially neuroprotective effect on the disease.[29,30] Studies in Experimental Autoimmune Encephalomyelitis (EAE) and *in vitro* observations in immunocompetent cells suggest that statins may exert immunomodulatory effects on both T cells and monocytes.[31,32] A small open-label study involving 30 RRMS patients who were receiving oral simvastatin demonstrated a decrease in the number and volume of Gd-enhancing lesions by 44% and 41%, respectively.[33] This was the first study to suggest a potential role for statins in the treatment of RRMS as determined by MRI measures.

Studies to evaluate the safety and efficacy of statins are currently in progress, but effects of add-on statin therapy with IFNs have raised concerns of a potentially incompatible, antagonistic effect of the combination. *In vitro* studies suggest that statins may block IFN-signaling pathways, thereby decreasing therapeutic efficacy.[34] In a double-blind, placebo-controlled trial of 24 patients randomized to receive atorvastatin (40 mg or 80 mg) in combination with subcutaneous IFN-beta 1a versus placebo, there was an increase in relapses and/or MRI-enhancing lesions in 10 of the 15 patients who were receiving

DOI:10.1002/MSJ

combination treatment versus only 1 of 9 in the placebo group.[35] In contrast, however, another trial that used only 20 mg of atorvastatin did not identify any differences with combination therapy.[36] Furthermore, interim analysis of the SIMCOMBIN trial investigating 80 mg of simvastatin or placebo plus IFN-beta 1a 30 mcg IM weekly also found no antagonistic effects of the IFN-statin combination.[37] A prior pilot study of 12 CIS patients randomized to placebo or simvastatin 80 mg daily plus IFN-beta 1a IM weekly found the combination to be safe and well tolerated.[38] The safety and efficacy of statin monotherapy in CIS is currently being investigated with results expected in the upcoming year.

Symptomatic Therapy–4-Aminopyridine/Fampridine

A very promising oral symptomatic therapy is Fampridine-SR, the sustained-release formulation of 4-aminopyridine (4-AP). Fampridine is unique among novel treatments as is optimizes functional reserve by enhancing conduction of damaged demyelinated axons via potassium channel blockade. In a Phase III randomized, double-blind, placebo-controlled trial of Fampridine, administered as adjunctive therapy, there was a significant improvement of 25% in walking speed seen after 14 weeks in 34.8% of the Fampridine group compared with 8.3% receiving placebo.[39] Walking speed was assessed by a timed 25-foot walk. Fampridine has a narrow efficacy-to-toxicity window with respect to risk of treatment-associated seizures. In this study, discontinuations due to adverse events occurred in 4.8% of the 229 treated patients and in none of the 72 patients in the placebo group. Three were considered serious, influenza, sepsis, and anxiety. A focal seizure, observed during sepsis, was considered possibly related to treatment. An additional 13 patients in the treatment group experienced various serious adverse events, none of which, however, were considered related to or led to discontinuation of treatment.

MONOCLONAL ANTIBODIES

Monoclonal antibodies are a group of selective agents, which bind to specific molecules on the surface of targeted cells, for example, T cells or B cells, both of which appear integral in the pathogenesis of MS. A number of mAbs currently approved or in clinical trials are considered among the most promising new therapies for MS. These include natalizumab, alemtuzumab, rituximab, and daclizumab. Although several of these agents have shown considerable promise as targeted immunotherapies, they have several limitations, including associated severe adverse events, infusion-related reactions, and development of neutralizing antibodies.[40]

Natalizumab

Natalizumab is a humanized mAb that targets alpha-4 integrin, an adhesion molecule on the surface of leukocytes, thereby interfering with entry of activated T cells into the CNS compartment. Reapproved in June 2006 by the US FDA and European authorities as monotherapy for the treatment of active relapsing forms of MS, natalizumab is reserved for patients who have failed to respond adequately to or cannot tolerate other conventional DMTs. To review, based on positive interim analyses of 2 Phase III trials, AFFIRM[41] and SENTINEL,[42] the FDA initially approved natalizumab in November 2004. Unfortunately, despite impressive efficacy results, natalizumab was unexpectedly quickly withdrawn from the market in February 2005, nearly 4 months after approval, after 2 patients who were receiving natalizumab in combination with IFN-beta 1a developed PML, an infection secondary to the reactivation of a commonly latent virus, the JC virus.[43] Additionally, a patient with Crohn disease treated with natalizumab, along with other immunosuppressants, also developed PML.[44] As such, an exhaustive investigation of all patients treated with natalizumab was conducted, and no additional cases of PML had been identified over a period of 18 months, leading to the drug's conditional reapproval. Moreover, a risk management plan to tighten PML surveillance was implemented, referred to as the Tysabri Outreach Unified Commitment to Health (TOUCH) program.[45]

Although it remains uncertain whether JC virus reactivation resulted from a broad immunosuppression secondary to combination therapy with IFN or as a specific effect of the mAb itself, natalizumab, nevertheless, serves as the prototypical lesson of the perils involved when novel, high-efficacy therapies with uncertain long-term safety profiles are introduced into a treatment paradigm. To date, there are more than 17,000 patients on natalizumab worldwide with no new cases of PML reported. Whereas this is extremely encouraging, it is important to emphasize that the time period of this safety data is still less than 2 years, the time period after which PML cases were identified. As such, it is still admittedly premature to determine the exact risk of developing PML-Informed discussions between clinicians and patients on natalizumab's as yet undefined risk-to-benefit ratio need to be emphasized when considering use of this agent in those failing to respond to conventional DMT.

Alemtuzumab

Alemtuzumab (CAMPATH-1H) is a humanized mAb that binds to the CD52 antigen on the surface of thymocytes, natural-killer (NK) cells, and B cells. It is FDA approved for the treatment of fludarabine-resistant chronic lymphocytic leukemia. Binding to the CD52 antigen results in antibody-dependent lysis and rapid removal of T cells from blood, bone marrow, and organs. This T-cell depletion lasts for an extended period of time, up to 16 months. A study by Coles and colleagues[46] assessed the use of alemtuzumab in early, very active MS and demonstrated a significant reduction in relapse rate of 91% as well as improvement in EDSS at 1 year. On the basis of these results, investigators designed the CAMMS223 trial. In this study, patients with RRMS were randomized to receive alemtuzumab or 44 mcg IFN-beta 1a subcutaneously 3 times a week. The trial was interrupted in 2005 because 3 unexpected cases of idiopathic thrombocytopenia purpura (ITP), including 2 fatalities occurred. In addition to ITP, autoimmune thyroiditis has also been identified in MS patients treated with alemtuzumab. After the suspension of alemtuzumab, 3 additional cases of ITP were identified early and treated successfully. A 2-year interim analysis presented at the 2007 American Academy of Neurology (AAN). Annual Meeting demonstrated that the alemtuzumab-treatment group had a 75% or greater reduction in risk of relapse ($P = .003$), and a 65% reduction in risk for a 6-month sustained accumulation of disability, when compared to IFN.[47] Results relating to 3 years were presented at European Committee for Treatment and Research in Multiple Sclerosis (ECTRIMS) in October 2007.[48] Analysis of the primary endpoint demonstrated the alemtuzumab-treated group experienced a 73% relapse reduction after 3 years when compared with patients taking IFN. Time to sustained accumulation of disability, another primary endpoint, showed that patients who were receiving alemtuzumab experienced at least a 70% reduction in risk for progression to clinically significant disability compared with those receiving IFN. Further results are expected to be presented at the 2008 AAN Meeting.

Despite its efficacy, potentially serious adverse events can occur months or years after dosing, specifically a 20% risk of autoimmune thyroiditis and a 2% risk of ITP.[49,50] With careful monitoring for both disorders in place, the study suggested that early diagnosis can lead to successful treatment. Two Phase III trials that compared alemtuzumab to IFN-beta 1a are currently recruiting, 1 in DMT-naïve patients and another in patients who failed to respond to an adequate trial of DMT.

Rituximab and B-Cell Involvement

There is a growing body of evidence that supports the role of B lymphocytes in the immunopathogenesis of MS. Although the exact mechanism remains unknown, several plausible possibilities include contribution to antibody-mediated inflammation-based myelin damage, encouragement of T cell-mediated tissue destruction, and a role in neurodegeneration.[51] As such, an agent that targets B cells offers a logical approach to MS therapy. Rituximab is a chimeric murine/human mAb that targets and selectively depletes CD20, an antigen present on pre-B and mature-B lymphocytes; however, it is not found on hematopoietc stem cells, plasma cells, or in normal tissues. The exact function of CD20 is not known. Rituximab is currently FDA approved for the treatment of non-Hogdkin lymphoma, refractory rheumatoid arthritis, and diffuse B-cell lymphoma.

Results of a Phase II randomized, placebo-controlled trial of 104 patients with RRMS who were receiving either rituximab 1000 mg intravenously 2 weeks apart versus placebo demonstrated a significant reduction of 91% in mean total Gd-enhancing lesions in the treatment group compared with placebo ($P < .0001$).[52] This reduction was achieved by Week 12 and sustained for the duration of the study. The number of new Gd-enhancing lesions and T2 lesion volume were also reduced. Moreover, there was a trend toward lower annualized relapse rates that reached significance by Week 24 but did not carry through to Week 48. Rituximab was well tolerated with similar mild to moderate adverse events reported in both groups, albeit a higher incidence of infusion reactions in the treatment group.

The presence of B-cell follicle-like structures in the cerebral meninges of 2 of 3 patients with secondary progressive MS suggests humoral autoimmunity may distinctly factor in progressive disease forms.[53] A preliminary study of rituximab in patients with primary progressive MS (PPMS), identified a limited effect on B cells in the cerebrospinal fluid.[54] A separate Phase II trial in PPMS is currently completed, and results should be available in 2008. Rituximab also appears to be a promising treatment option for patients with neuromyelitis optica (NMO), a humorally mediated demyelinating disease with preference for the spinal cord and optic nerves. In an open-label study by Cree and colleagues, there was a decrease in relapse rates and improvement in disability scores in 8 patients with NMO who had failed to respond to prior treatment with other immunosuppressant regimens.[55]

In 2007, the FDA reported 2 cases of 10,000 patients with systemic lupus erythematosus (SLE) who had received rituximab and had developed PML.[56] Of note, these patients had also been treated previously with other immunosuppressants, and PML, itself, has been reported in the SLE population in patients who have not been treated with rituximab. As such, a causal relation between PML and rituximab treatment cannot be established from this report. Nevertheless, clinicians and patients should be aware of this risk when they are considering rituximab as a treatment option, particularly for those patients with a prior history of immunosuppressant therapy.

Daclizumab

Daclizumab, an FDA-approved treatment for the prevention of graft-versus-host disease in renal transplant patients, is a humanized mAb that binds to CD25, the alpha-chain of interleukin (IL)-2 receptor, resulting in decreased IL-2-mediated stimulation of T-cell function. Preliminary studies have evaluated the effect of adding daclizumab to IFN-beta therapy. In an open-label trial of 10 patients with MS who had high clinical and MRI activity, there was a 78% reduction in new Gd-enhancing lesions.[57] Montalban and colleagues presented results of a recent Phase II study of daclizumab add-on therapy at the 2007 ECTRIMS meeting. In this study, the CHOICE trial, 230 actively relapsing MS patients on IFN-beta were randomized to either daclizumab (1 or 2 mg/kg subcutaneously) every 2 weeks for 24 weeks or placebo. There was a significant reduction of 72% in mean number of new or enlarging Gd-enhancing lesions at Weeks 8–24 in the 2-mg/kg-treatment arm compared with placebo ($P = .004$). In addition, a 25% reduction in Gd-enhancing lesions was seen on the lower dose, albeit not statistically significant. Moreover, a nonsignificant reduction of 33% in relapse rate was noted in the high-dose group; however, the study was not designed to compare relapse rates. Adverse effects included infections, urinary tract infections, and cutaneous events. Serious infections were more common in the high-dose treatment group.[58]

ANTIGEN-BASED IMMUNE THERAPY (MBP8298) AND DNA VACCINATION (BHT-3009)

Antigen-based immune therapy is an attractive therapeutic approach to the treatment of autoimmune disorders. Tolerance can be induced to specific antigens by oral application of, for example, myelin basic protein (MBP) resulting in production of antigen-specific regulatory T cells in the periphery. This triggers development of bystander suppression after antigen recognition by the target organ. Several studies, both experimental and clinical, have been conducted in an attempt to implement antigen-specific therapy, albeit with limited clinical success in humans. For instance, trials of oral myelin[59] and altered peptide ligands derived from immunodominant peptides of MBP were disappointing.[60] Of late, an approach that has shown some promise involves repetitive application of MBP8298, an MBP-derived peptide with residues representing the minimal epitope necessary for antibody response in the cerebrospinal fluid (CSF) of MS patients. Based on data from an open-label study, which investigated MBP8298 in chronic, progressive forms of MS, 500 mg of MBP8298 induced lasting immunological tolerance, interestingly in a subclass of patients positive for specific HLA-DR haplotypes DR2, DR4 and DR7.[61] A Phase II trial in chronic-progressive disease demonstrated that fewer patients in the treatment group progressed, as determined by EDSS scores. Compared with placebo, the effect was most impressive in the subgroup HLA-DR2 and/or DR4 individuals.[62] A 2-year Phase II/III study that investigated MBP8298 in SPMS patients with specific HLA-DR2/4 haplotypes is currently recruiting.

DNA-derived immunization could also potentially augment the efficacy of antigen-induced tolerance. An example is BHT-3009, a DNA-plasmid vaccine that encodes full-length MBP. In a Phase II trial lead by Garren and colleagues,[63] 289 patients were randomized to placebo, 0.5 mg BHT-3009 IM or 1.5 mg BHT-3009. Median rate of development of Gd-enhancing lesions between 24 and 28 weeks was decreased by 50% in the 0.5 mg group, with no significant difference; however, it was noted in the 1.5 mg group. Additionally, no significant effect was noted on relapse rate or T2-lesion burden.

REMYELINATION/REPAIR AND NEUROPROTECTION

The last decade has seen several advances in the understanding of mechanisms responsible for oligodendrogenesis and myelination with several plausible translational targets identified. Whereas clinical application of such strategies is likely decades away, unraveling the complexities of remyelination is critical to our understanding of repair of tissue damage and recovery of function. In addition to laboratory efforts to demystify remyelination, a focus on neuroprotective agents is an area of great interest in identifying methods to maintain axonal integrity and promote regeneration.

INDIVIDUALIZING THERAPIES

As the therapeutic horizon continues to change with the advent of new options, choice of first-line therapy will undoubtedly become a more complicated decision. Natalizumab has certainly taught us caution when using short-term safety data and when experiencing unexpected adverse events in exchange for potentially better outcomes. Without question, the wide spectrum and unpredictable nature of MS itself limits our ability to prognosticate well, and clinicians continue to face a trial-and-error treatment period, which potentially threatens a critical therapeutic interval in the course of our patients' disease. Several new modalities, which aim to identify treatment response, are currently being investigated. Stratifying therapeutic subgroups by using biomarkers could foster the development and application of individualized therapies. Knowing which therapy will work best early on would be paramount. In addition to genetics, area of great promise involves the field of proteomics. Proteomics is the study of patterns of serum proteins from individual patients. Such fingerprint profiles could be instrumental to the discovery of new biomarkers, which indicate not only a particular disease type but also successful response to specific therapies.

CONCLUSION

In summary, the field of MS research offers insight into the challenges of developing neurotherapeutics for chronic neurologic diseases, particularly those with complex pathophysiology and unpredictable clinical spectra. To date, a plethora of novel agents in various clinical trials are currently being tested in MS with the hope of broadening the presently limited therapeutic arsenal. These include agents that focus on optimizing disease modification as well as neuroprotection and tissue repair. Although several current and future investigational therapies appear promising, it is clear that a greater understanding of MS disease pathophysiology is still needed. As neuropathologic studies have demonstrated, patterns of disease are as variable, if not more heterogeneous, than clinical subtypes.[64] It remains to be seen whether therapies under clinical investigation will replace the available armamentarium of DMTs or find designation among a bevy of individualized therapies specific to disease subtype and/or treatment response.

REFERENCES

1. Raine CS, Scheinberg LC. On the immunopathology of plaque development and repair in multiple sclerosis. *J Neuroimmunol* 1988; 20: 189–201.
2. Trapp BD, Peter J, Ransohoff RM, et al. Axonal transaction in the lesions of multiple sclerosis. *N Engl J Med* 1998; 338: 278–285.
3. Bitsch A, Schuchardt J, Bunkowski S, et al. Acute axonal injury in multiple sclerosis. Correlation with demyelination and inflammation. *Brain* 2000; 123: 1174–1183.
4. Johnson KP, Brooks BR, Cohen J, et al, Copolymer 1 Multiple Sclerosis Study Group. Extended use of glatiramer acetate (Copaxone) is well tolerated and maintains its clinical effect on multiple sclerosis relapse rate and degree of disability. *Neurology* 1998; 50(3): 701–708.
5. Randomized double-blind placebo-controlled study of interferon beta-1a in relapsing-remitting multiple sclerosis. PRISMS (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group. *Lancet* 1998; 352: 1498–1504.
6. PRISMS Study Group; University of British Columbia MS/MRI Analysis Group. PRISMS-4: Long-term efficacy of interferon-beta-1a in relapsing MS. *Neurology* 2001; 56(12): 1628–1636.
7. Comi G, Filippi M, Barkhof F, et al, Early Treatment of Multiple Sclerosis Study Group. Effect of early interferon treatment on the conversion to definite multiple sclerosis; a randomized study. *Lancet* 2001a; 357(9268): 1576–1582.
8. Jacobs LD, Beck RW, Simon JH, et al. Intramuscular interferon beta 1a therapy initiated during a first demyelinating vent in multiple sclerosis. CHAMPS Study Group. *N Engl J Med* 2000; 343(13): 898–904.
9. Kinkel RP, Killman C, O'Connor P, et al, CHAMPIONS Study Group. IM interferon beta-1a delays definite multiple sclerosis 5 years after a first demyelinating event. *Neurology* 2006; 66(5): 678–684.
10. Kappos L, Polman CH, Freedman MS, et al. Treatment with interferon beta-1b delays conversion to clinically definite and McDonald MS in patients with clinically isolated syndromes. *Neurology* 2006b; 67(7): 1242–1249.
11. Johnson KP. Control of multiple sclerosis relapses with immunomodulatory agents. *J Neurol Sci* 2007; 256(suppl. 1): S23–S28.
12. Amato MP, DeStefano N. Longitudinal followup of "benign" multiple sclerosis at 20 years. *Neurology* 2007; 68(7): 496–500.

13. Mikol DD, Barkhof F, Chang PK, et al. The REGARD trial: a randomized assessor-blinded trial comparing interferon beta 1a and glatiramer acetate in relapsing-remitting multiple sclerosis. *Mult Scler* 2007; 13(suppl. 2): S269, Abstract 119.
14. Marrie RA, Cutter G, Tyry T, et al. Changes in the ascertainment of multiple sclerosis. *Neurology* 2005; 65: 1066–1070.
15. Wolansky L, Cook S, Skurnick J, et al. Betaseron [IFNB-1b] vs. Copaxone [glatiramer acetate] in MS with triple-dose gadolinium and 3 T MRI endpoints (BECOME): announcement of final primary study outcomes. *Mult Scler* 2007; 12(suppl. 2): S58, Poster 206.
16. Cadavid D, Wolansky L, Cook S, et al. Betaseron [IFNB-1b] vs. Copaxone [glatiramer acetate] in MS with triple-dose gadolinium and 3 T MRI endpoints (BECOME): announcement of secondary clinical outcomes. *Mult Scler* 2007; 12(suppl. 2): S58, Poster 207.
17. The BEYOND (Betaseron Efficacy Yielding Outcomes of a New Dose Study) http://www.reuters.com/article/companyNewsAndPR/idUSL2923113720071029 (Accessed January 2008).
18. Lublin F, Cutter G, Elfont R, et al. A trial to assess the safety of combining therapy with interferon beta-1a and glatiramer acetate in patients with relapsing MS. *Neurology* 2001; 56(suppl): A148.
19. Lublin FD, Baier M, Cutter G, et al. Results of the extension of a trial to assess the longer term safety of combining interferon beta-1a and glatiramer acetate. *Neurology* 2002; 58(suppl. 3): A85.
20. Kappos L, Antel J, Comi G, et al. Oral fingolimod (FTY720) for relapsing multiple sclerosis. *N Engl J Med* 2006a; 355(11): 1124–1140.
21. Barske C, Osinde M, Mir AK, et al. FTY720 (fingolimod) enhances the number of immature and mature oligodendrocytes. Program and abstracts of the ECTRIMS 2007: 23rd Congress of the European Committee for Treatment and Research in Multiple Sclerosis October 11–14, 2007; Prague, Czech Republic, Poster 498.
22. Miron VE, Jung CG, Kim HJ, et al. FTY720 modulates human oligodendrocyte progenitor process extension and survival. *Ann Neurol* 2008; 63(1): 61–71.
23. Gold R, Fox R, Dawson K, et al. Two phase 3 studies to determine the efficacy and safety of BG00012, a novel, oral fumaric acid derivative, in patients with relapsing multiple sclerosis. *Mult Scler* 2007; 13(suppl. 2): S173.
24. Vandermeeren M, Janssens S, Borgers M, Geysen J. Dimethylfumarate is an inhibitor of cytokine-induced E-selectin, VCAM-1, and ICAM-1 expression in human endothelial cells. *Biochem Biophys Res Commun* 1997; 234: 19–23.
25. de Jong R, Bezemer AC, Zomerdijk TP, et al. Selective stimulation of T helper 2 cytokine responses by anti-psoriasis agent monomethylfumarate. *Eur J Immunol* 1996; 26: 2067–2074.
26. Kappos L, Miller DH, MacManus DG, et al. Efficacy of a novel oral single-agent fumarate, BG00012, in patients with relapsing-remitting multiple sclerosis: results of a phase 2 study. *J Neurol* 2006; 253: 27.
27. Leist TP, Vermerch P. The potential role for cladribine in the treatment of multiple sclerosis: clinical experience and development of an oral tablet formulation. *Curr Med Res Opin* 2007; 23(11): 2667–2676.
28. O'Connor PW, Li D, Freedman MS, et al. A phase II study of the safety and efficacy of teriflunomide in multiple sclerosis with relapses. *Neurology* 2006; 66(6): 894–900.
29. Stuve O, Prod'homme T, Youssef S, et al. Statins as potential therapeutic agents in multiple sclerosis. *Curr Neurol Neurosci Rep* 2004; 4: 237–244.
30. Neuhaus O, Arcgelos JJ, Hartung HP. Statins in multiple sclerosis: a new therapeutic option? *Mult Scler* 2003; 9: 429–430.
31. Greenwood J, Walters CE, Pryce G, et al. Lovastatin inhibits brain endothelial cell Rho-mediated lymphocyte migration and attenuates experiemental auto-immune encephalomyelitis. *FASEB J* 2003; 17: 905–907.
32. Youssef S, Stuve O, Patarroyo JC, et al. The HMG-CoA reductase inhibitor atorvastatin, promotes loss of therapeutic effect of IFN-beta in multiple sclerosis. *Neurology* 2007; 68(suppl. 1): A364.
33. Vollmer T, Key L, Durkalski V, et al. Oral simvastatin treatment in relapsing remitting multiple sclerosis. *Lancet* 2004; 363: 1607–1608.
34. Dhawan N, Reder A. Statins block interferon signaling in human immune cells: potential loss of the therapeutic effect of IFN-beta in multiple sclerosis. *Neurology* 2007; 68(suppl. 1): A364.
35. Birnbaum G, Altafullah I, Reder A. A double blind, placebo controlled trial of atorvastatin in combination with subcutaneous interferon beta 1a in persons with multiple sclerosis. *Neurology* 2007; 68(suppl. 1): A206–A207.
36. Orefice G, Quarantelli M, Salvatore P, et al. An open label randomised clinical trial of atorvastatin in combination with IFN beta-1a in active MS patients: 12 months results. Program and abstracts of the ECTRIMS 2007: 23rd Congress of the European Committee for Treatment and Research in Multiple Sclerosis October 11–14, 2007; Prague, Czech Republic, Poster 202.
37. Sorensen PS, Frederiksen JL, Lycke L, et al. Does simvastatin antagonize the effect of interferon-beta? Interim safety analysis of ongoing SIMCOMBIN study. Program and abstracts of the ECTRIMS 2007: 23rd Congress of the European Committee for Treatment and Research in Multiple Sclerosis October 11–14, 2007; Prague, Czech Republic, Abstract 85.
38. Markovic-Plese S, Speer D, Jin J, et al. Statin and intramuscular interferon beta-1a combination therapy is safe and well tolerated in patients with clinical isolated syndrome suggestive of multiple sclerosis, a pilot study. Program and abstracts of the ECTRIMS 2007:23rd Congress of the European Committee for Treatment and Research in Multiple Sclerosis; October 11–14; Prague, Czech republic, Poster 232a.
39. Goodman A, Schwid S, Brown T, et al. A Phase 3, Mult-Center Trial of Oral, Sustained-Release Fampridine (4-Aminpyridine) in Multiple Sclerosis. AAN Meeting Abstract 2007[S32.003].
40. Hohlfeld R, Wekerle H. Drug Insight: using monoclonal antibodies to treat multiple sclerosis: a randomized, double-blind, placebo-controlled trial. *Nat Clin Pract Neurol* 2005; 1: 34–44.
41. Polman CH, O'Connor PW, Havrdova E, et al. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. *N Engl J Med* 2006; 354: 899–910.

42. Rudick RA, Stuart WH, Calabresi PA, et al. Natalizumab plus interferon beta-1a for relapsing multiple sclerosis. *N Engl J Med* 2006; 354: 911–923.
43. Kleinschmidt-DeMasters BK, Tyler KL. Progressive multifocal leukoencephalopathy complicating treatment with natalizumab and interferon beta-1a for multiple sclerosis. *N Engl J Med* 2005; 353: 369–374.
44. Langer-Gould A, Atlas SW, Green AJ, et al. Progressive multifocal leukoencephalopathy after natalizumab therapy for Crohn's disease. *N Engl J Med* 2005; 353: 362–368.
45. Department of Health and Human Services. *Tysabri risk Minimization Action Plan: Summary of TOUCH.* Rockville, MD: Deparment of Health and Human Services, 2007; http://www.fda.gov/cder/drug/infopage/natalizumab/RiskMAP.pdf
46. Coles AJ, Cox A, Le Page E, et al. The window of therapeutic opportunity in multiple sclerosis : evidence from monoclonal antibody therapy. *J Neurol* 2006; 253(1): 98–108.
47. Coles AJ, The CAMMS223 Study Group. Efficacy of alemtuzumab in treatment-naïve relapsing-remitting multiple sclerosis; analysis after two years of study CAMMS223 [AAN abstract S12.004]. *Neurology* 2007; 68(suppl. 1): A100.
48. Coles AJ, On behalf of the CAMMS223 Study Group. Alemtuzumab improved multiple sclerosis functional composite scores and delayed time to first relapse at 2-year interim analysis compared to subcutaneous interferon-beta-1a. *Mult Scler* 2007; 13(suppl. 2): S166, (abstract).
49. Coles AJ, Wing M, Smith S, et al. Pulsed monoclonal antibody treatment and autoimmune thyroid disease in multiple sclerosis. *Lancet* 1999a; 354(9191): 1691–1695.
50. Sullivan H, The CAMMS223 Study Group. ITP following treatment of multiple sclerosis patients with alemtuzumab in CAMMS223; case reports and risk management plan implementation [AAN abstract S32.004]. *Neurology* 2007; 68(supp. 1): A207.
51. Nikbin B, Bonab MM, Khosravi F, Talebian F. Role of B cell follicles with germinal centers in the meninges of patients with relapsing remitting multiple sclerosis (RRMS) [AAN abstract S12.003]. *Neurology* 2007; 68(suppl. 1): A99.
52. Hauser S, Waubaut E, Arnold D, et al. A phase II randomized, placebo-controlled, multi-center trial of rituximab in adults with relapsing remitting multiple sclerosis (RRMS) [AAN abstract S12.003]. *Neurology* 2007; 68(suppl. 1): A99.
53. Serafini B, Rosicarelli B, Magliozzi R, et al. Detection of ectoptic B-cell follicles with germinal centers in the meninges of patients with secondary progressive multiple sclerosis. *Brain Pathol* 2004; 14: 164–174.
54. Monson NL, Cravens PD, Frohman EM, et al. Effect of rituximab on the peripheral blood and cerebrospinal fluid B cells in patients with primary progressive multiple sclerosis. *Arch Neurol* 2005; 62: 258–264.
55. Cree BA, Lamb S, Morgan K, et al. An open label study of the effects of rituximab in neuromyelitis optica. *Neurology* 2005; 64: 1270–1272.
56. US Food and Drug Administration Center for Drug Evaluation and Research. *Rituximab (Marketed and Rituxan) Information.* Rockville, MD Department of Health and Human Services, 2006.
57. Bielekova B, Richert N, Howard T, et al. Humanized anti-CD25 (daclizumab) inhibits disease activity in multiple sclerosis patients failing to respond to interferon beta. *Proc natl Acad Sci U S A* 2004; 101(23): 8705–8708, Epub 2004 May 25. Erratum in: Proc Natl Acad Sci U S A 2004; 101(50): 17565.
58. Montalban X, Wynn D, Kaufman M, et al. Preliminary CHOICE results: a phase 2, randomised, double-blind, placebo-controlled multicentre study of subcutaneous daclizumab in patients with active, relapsing forms of multiple sclerosis on interferon beta. Program and abstracts of the ECTRIMS 2007: 23rd Congress of the European Committee for Treatment and Research in Multiple Sclerosis; October 11–14, 2007; Prague, Czech Republic, Abstract 50.
59. Panitch H, Francis G, The Oral Myelin Study Group. Clinical results of a phase III trial of oral myelin in relapsing-remitting multiple sclerosis. *Ann Neurol* 1997; 42: 459.
60. Bielekova B, Goodwin B, Richert N, et al. Encephalitogenic potential of the myelin basic protein peptide (amino acids 83–99) in multiple sclerosis: results of a phase II clinical trial with altered peptide ligand. *Nat Med* 2000; 6: 1167–1175.
61. Warren KG, Catz I, Wucherpfennig KW. Tolerance induction to myelin basic protein by intravenous synthetic peptides containing epitope P85 VVHFFKNIVTP96 in chronic progressive multiple sclerosis. *J Neurol Sci* 1997; 152: 31–38.
62. Warren KG, Catz I, Ferenczi LZ, Krantz MJ. Intravenous syndrome peptide MBP8298 delayed disease progression in an HLA Class II-defined cohort of patients with progressive multiple sclerosis: results of a 24-month double-blind placebo-controlled clinical trial and 5 years of follow-up treatment. *Eur J Neurol* 2006; 13: 887–895.
63. Garren H, Krasulova E, Havrdova E, et al. Phase IIb trial of a MBP encoding DNA plasmid (BHT-3009) for the treatment of relapsing-remitting multiple sclerosis. Program and abstracts of the ECTRIMS 2007: 23rd Congress of the European Committee for Treatment and Research in Multiple Sclerosis October 11–14, 2007; Prague, Czech Republic, Abstract 48.
64. Lucchinetti C, Bruck W, Parisi J, et al. Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. *Ann Neurol* 2000; 47: 707–717.

PRISMATIC CASE

Report of Fertility in Woman with Predominantly 46,XY Karyotype in Family with Multiple Disorders of Sexual Development: Review of Prismatic Case

Miroslav Dumic,[1] Karen Lin-Su,[2] Natasha I. Leibel,[3] Srecko Ciglar,[4]
Giovanna Vinci,[5] Ruzica Lasan,[1] Saroj Nimkarn,[2] Jean D. Wilson,[6]
Ken McElreavey,[5] and Maria I. New[2]

[1] Department of Pediatric Endocrinology and Diabetes, University Hospital Rebro, Zagreb, Croatia
[2]Department of Pediatric Endocrinology/Division of Adrenal Steroid Disorders, Mount Sinai School of Medicine, New York, NY
[3]Department of Pediatric Endocrinology, Columbia University, New York, NY
[4] Department of Gynecology and Obstetrics, Clinical Hospital Merkur, Zagreb, Croatia
[5] Reproduction, Fertility, and Populations, Institut Pasteur, Paris, France
[6] Department of Internal Medicine/Endocrinology,
University of Texas Southwestern Medical School, Dallas, TX

We report a prismatic case of a 46,XY woman who gave birth to a daughter.[1] This fertile woman was investigated with molecular genetic analyses for the genes encoding the testes, and the genes were found to be normal for a male. The patient menstruated normally until menopause. She had 2 pregnancies: one was spontaneously aborted, and the other was a 46,XY female with normal female genitalia whose diagnosis was pure gonadal dysgenesis. The Y in the daughter was inherited from the father. The 46,XY mother has a normal female phenotype. Postmenopausal ovariectomy was performed because of the risk of an ovarian gonadoblastoma in a 46,XY female. The 46,XY karyotype of the ovary is unprecedented (see Figure 1). The pedigree reveals many sibs of the 46,XY mother to have abnormalities of sexual development. Although this is a historic case of fertility in a 46,XY female, we do not know the prevalence of fertile 46,XY females, as mothers are not routinely karyotyped.

REFERENCE

1. Dumic M, Lin-Su K, Leibel NI, et al. Report of fertility in a woman with a predominantly 46,XY karyotype in a family with multiple disorders of sexual development. *J Clin Endocrinol Metab* 2008; 93: 182–189.

Address Correspondence to:

Maria I. New
Mount Sinai School of Medicine
New York, NY
Email: maria.new@mssm.edu

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20046

© 2008 Mount Sinai School of Medicine

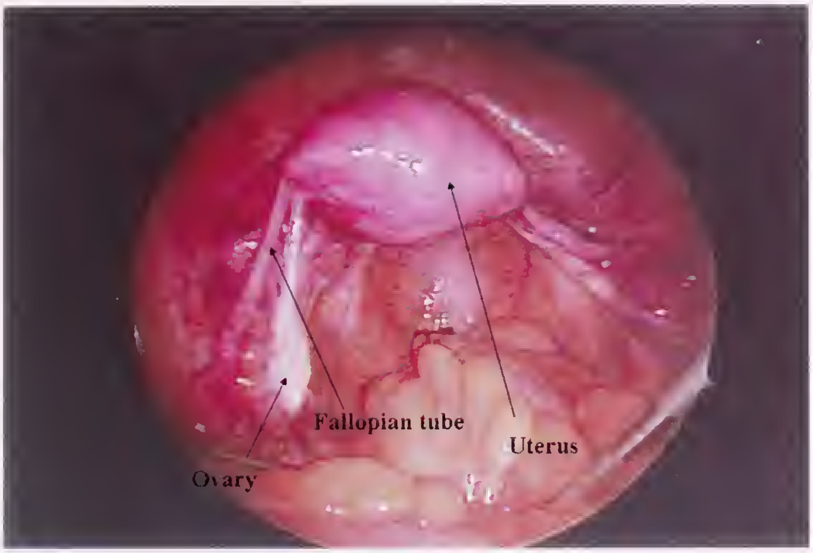


Fig 1. The ovary of a 46,XY female, the karyotype of which was 93% 46,XY and 6% 45,X. The ovary was sex-determining region Y (SRY)–positive. The female menstruated regularly and had 2 pregnancies. The woman gave birth to a female with 46,XY gonadal dysgenesis in whom the Y was transmitted from the father. Molecular analysis of the mother and daughter revealed coding sequences of 9 genes known to be involved in sexual development [sex-determining region Y box 9 (*SOX9*); splicing factor 1 (*SF1*); doublesex and mab-3 related transcription factor 1 (*DMRT1*); doublesex and mab-3 related transcription factor 3 (*DMRT3*); testis-specific protein, Y-linked (*TSPYL*); blepharophimosis, epicanthus inversus and ptosis, candidate 1 (*BPESC1*); desert hedgehog homolog (*DHH*); wingless-type MMTV integration site family, member 4 (*WNT4*); *SRY*; and dosage-sensitive sex reversal on chromosome X, gene 1 (*DAX1*)] that are identical to the normal reference sequences. This is the first report of a fertile 46,XY human female, although this has been reported in the mouse and horse.

DOI:10.1002/MSJ

MOUNT SINAI JOURNAL OF MEDICINE

VOL. 75, NO. 3
MAY/JUNE 2008

Issue Theme
Anxiety and Depression: Leading Edge of Therapy

A Journal of Translational and Personalized Medicine

Introduction by Dennis Charney, MD, Dean, and Sanjay Mathew, MD, Asst. Professor of Psychiatry, Mount Sinai School of Medicine

Obsessive Compulsive Disorder: Translational Approaches and New Somatic Developments

Major Depression: Emerging Pharmacotherapeutics

Bipolar Disorder: Putative Drugs and Targets

Anxiety Disorder: Comprehensive Review of Pharmacotherapies

Treatment-Resistant Depression: Neurostimulatory Therapeutics

Treating Pediatric Patients with Antripsychotic Drugs: Balancing Benefits and Safety

Special Clinical Features

Predictive Genetic Testing in Minors for Adult-onset Genetic Diseases
Thyroid Nodules: Update on Differential Diagnosis and Management
CCR5 Inhibitors: New Class of Drugs for AIDS

WILEY-BLACKWELL

Discover
Something
Great

Distinctive Content

Content
Journals • Backfile Collections
Reference Works & Encyclopedias
OnlineBooks™
Current Protocols Lab Manuals
Evidence-Based Resources

Access
Single article/chapter or full title
One-time fee, subscription,
or Pay-Per-View options
Unlimited concurrent usage

Capabilities
E-mail Alerts • EarlyView®
Citation download & tracking
Remote access • Mobile Edition®

Promote the process of scientific discovery with a different breed of research system: **Wiley InterScience**®. Long distinguished by its scope and extensive coverage of emerging disciplines, Wiley's authoritative content spans the spectrum of scientific, medical, technical, and business specialties. The dynamic **Wiley InterScience** platform adds value to your research by providing simple desktop access to this preeminent digital library of journals, reference works, books, laboratory manuals, evidence-based databases, and more. Researchers reap the combined benefits of top-tier resources and up-to-the-minute functionality. Experience the difference **Wiley InterScience** can make at your institution today.

WILEY
InterScience®
DISCOVER SOMETHING GREAT

WILEY
1807-2007 KNOWLEDGE FOR GENERATIONS

www.interscience.wiley.com

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

RECEIVED
AUG 28 2008
Levy Library

Editor-in-Chief

Penny A. Asbell

Deputy Editors

Dennis Charney
Maria New
C. Warren Olanow

Associate Editors

Stuart Aaronson
David H. Adams
Erwin Bottinger
Eric Hollander
Ravi Iyengar
Mary Klotman
Paul Klotman
Ihor R. Lemischka
Fred Lublin
Michael L. Marin
Thomas G. McGinn
Ira Nash
Rosamond Rhodes
Hugh Sampson
Mary Sano
Anthony Squire
Fred Suchy
Savio Woo

Editorial Board

William J. Bremner	University of Washington
Thomas M. Coffman	Duke University
Dwight Evans	University of Pennsylvania
Gabriel Haddad	University of California, San Diego
Heidi E. Hamm	Vanderbilt University
Matti Häyry	University of Manchester
Eve J. Higginbotham	Morehouse School of Medicine
Katherine A. High	University of Pennsylvania
Tony Hope	University of Oxford
John H. Krystal	Yale University
Raju Kucherlapati	Harvard Medical School
Robert J. Lefkowitz	Duke University
Robert Lisak	Wayne State University
George Lister	University of Texas
Philip A. Marsden	University of Toronto
Charles Nemeroff	Emory University
Scott Rauch	Harvard Medical School
Bernard Roizman	University of Chicago
Alan F. Schatzberg	Stanford University
Inder Verma	The Salk Institute
Jeffrey A. Whitsett	University of Cincinnati
Allan Wolkoff	Albert Einstein College of Medicine
Noam Zohar	Bar-Ilan University

Guest Co-editors

Sanjay J. Mathew, MD
Assistant Professor of Psychiatry
Mood and Anxiety Disorders Program
Mount Sinai School of Medicine

Dennis S. Charney, MD
Dean and Professor of Psychiatry, Neuroscience,
Pharmacology and Systems Therapeutics
Mount Sinai School of Medicine

Managing Editor

Stephen Cohen
stephen.cohen@mssm.edu
212-241-6304

Publisher

John Wiley & Sons
Anna Wesley, Associate Editor
awesley@wiley.com

"The Scream" by Edvard Munch

MOUNT SINAI JOURNAL OF MEDICINE
Issue Theme
Anxiety and Depression: Leading Edge of Therapy
A Journal of Translational and Personalized Medicine
Special Clinical Features
WILEY-BLACKWELL

Painted and revised by the Norwegian artist during the period from 1893 to 1910, "The Scream" became one of the most widely recognized emblems of the 20th century's "Age of Anxiety." © 2008 The Munch Museum/The Munch-Ellingsen Group/Artists Rights Society (ARS), New York.

Themes for future issues of *MSJM*, each shown with Guest Editor(s)

- Diabetes type 1 (Ronald Tamler)
- Hospital medicine (Andrew Dunn)
- Advanced lung disease (Maria Padilla)
- Personalized medicine (Erwin Bottinger)
- Zoonotic infections (Stephanie Factor and Mary Klotman)
- Alzheimer's disease (Nilos Robakis)
- Traumatic Brain Injury (Andrew Jagoda)
- Medical education (Lisa Coplit)
- Drug discovery (Ravi Iyengar)
- Obstetrics/gynecology and reproductive services (Rhoda Sperling)

©2008 Mount Sinai School of Medicine. All rights reserved. No part of this publication may be reproduced in any form or by any means, except as permitted under section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the publisher, or authorization through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, Phone: (978) 750-8400; Fax: (978) 750-4770. Such requests and other permission inquiries should be addressed to the Permissions Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; Phone: (201) 748-6011; Fax: (201) 748-6008; E-mail: permreq@wiley.com.

Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine (Print ISSN: 0027-2507; Online ISSN: 1931-7581 at Wiley InterScience, www.interscience.wiley.com) is published twelve times per year for the Mount Sinai School of Medicine by Wiley Subscription Services, a Wiley Company, 111 River Street, Hoboken, NJ 07030.

Send subscription inquiries to: Subscription Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-6645.

Offprint sales and inquiries should be addressed to the Reprint Billing Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-8776. All other inquiries should be addressed to the Customer Service Department at (201) 748-6645.

Individual subscription price: $95.00 in North America, $120.00 outside North America. Payment must be made in US dollars drawn on a US bank. Claims for undelivered copies will be accepted only after the following issue has been delivered. For more subscription prices and information, please visit www.mountsinaijournal.org.

Change of address: Please forward six weeks prior to move: present mailing label with change of address. Please enclose a copy of the mailing label or cite your subscriber reference number in order to expedite handling. Missing copies will be supplied when losses have been sustained in transit and where reserve stock permits. Send claims in care of: *Mount Sinai Journal of Medicine*, c/o John Wiley & Sons, Inc., Attn: Journals Admin Dept. UK, 111 River Street, Hoboken, NJ 07030.

Cancellations: Subscription cancellations will not be accepted after the first issue has been mailed. Periodicals postage paid at Hoboken, NJ and at additional mailing offices.

Postmaster: Send address changes to *Mount Sinai Journal of Medicine*, Subscription Distribution, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Printed in the United States of America.

Advertising inquiries should be addressed to: Advertising Sales Manager, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Phone: (201) 748-6921; Fax: (201) 748-6207. Advertising in this journal is published with a caveat emptor understanding. The Publisher, Editors, and members of the Editorial Board do not imply warranty of products, nor quality, validity, or approval of education material offered by such advertising.

Indexed by: Biological Abstracts® (Thomson ISI), Chemical Abstracts Service/SciFinder (ACS), Current Contents®/Clinical Medicine (Thomson ISI), EMBASE/Excerpta Medica (Elsevier), Index Medicus/MEDLINE/PubMed (NLM), Science Citation Index Expanded™ (Thomson ISI), SCOPUS (Elsevier), Web of Science® (Thomson ISI).

This journal is printed on acid-free paper.

Disclaimer: All articles published, including but not limited to original research, reviews, clinical notes, clinical pearls, prismatic cases, clinician's corner, editorials, reports, and letters, represent the opinions and views of the authors and do not reflect any official policy or medical opinion of the Mount Sinai School of Medicine, the institutions with which the authors are affiliated, or the Publisher unless this is clearly specified. Articles published herein should not be relied upon as recommending or promoting a specific method, diagnosis, or treatment by physicians for any particular patient or disorder. Although the Editors and Publisher believe that drug selections and dosages and the specifications and usage of equipment and devices as set forth herein are in accord with current recommendations and practice at the time of publication, they accept no legal responsibility for any errors or omissions, and make no warranty, express or implied, with respect to material contained herein.

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

CONTENTS
VOL 75, NO 3

Theme Introduction

171 Leading Edge of Pharmacotherapy in Serious Mental Illness
Sanjay J. Mathew, Dennis S. Charney

Theme Articles

174 Obsessive Compulsive Disorder: Translational Approaches and New Somatic Developments
Heather A. Berlin, Eric Hollander

204 Major Depression: Emerging Pharmacotherapeutics
Srijan Sen, Gerard Sanacora

226 Bipolar Disorder: Putative Drugs and Targets
Carlos A. Zarate, Husseini K. Manji

248 Anxiety Disorder: Comprehensive Review of Pharmacotherapies
Ellen J. Hoffman, Sanjay J. Mathew

263 Treatment-resistant Depression: Neurostimulatory Therapeutics with Focus on Deep Brain Stimulation
Dani Dumitriu, Kate Collins, Ron Alterman, Sanjay J. Mathew

276 Treating Pediatric Patients with Antipsychotic Drugs: Balancing Benefits and Safety
Iliyan Ivanov, Alexander Charney

Feature Departments

287 Controversy
Predictive Genetic Testing in Minors for Adult-Onset Genetic Diseases
Pascal Borry, Tom Goffin, Herman Nys, Kris Dierickx

297 New Drugs and the FDA (Editor: Jeffrey Newcorn)
CCR5 Entry Inhibitors: New Class of Drugs for AIDS
Andrew Jacovina, Maria I. New

Special Feature

299 Thyroid Nodules: A Clinical Guide Evaluation and Treatment
Juan C. Galofré, Stylianos Lomvardias, Terry F. Davies

THEME INTRODUCTION

Anxiety and Depression: Leading Edge of Therapy

Sanjay Mathew, MD and Dennis Charney, MD

Mount Sinai School of Medicine
New York, NY

We are pleased to introduce a special issue of the *Mount Sinai Journal of Medicine* on the therapeutics of psychiatric disorders. Serious psychiatric disorders, comprising unipolar and bipolar disorders, anxiety disorders, and schizophrenia, among many other conditions, pose formidable challenges for clinicians in all specialties. Despite progress in psychiatric therapeutics over the past several decades, significant gaps remain in both the effectiveness of common treatments and their translation into clinical practice. Results from the National Comorbidity Survey Replication, a nationally representative household survey, illustrate some of these challenges. Severe impairment in social and occupational functioning was reported by a significantly higher proportion of persons with mental disorders (42.0%) in comparison with persons with chronic medical disorders (24.4%);[1] however, treatment was provided for a significantly lower proportion of mental disorders (21.4%) in comparison with chronic medical disorders (58.2%). The proportion of psychiatric patients receiving appropriate, evidence-based treatment is even lower.

The 6 articles in this issue, focusing on pharmacological management and experimental therapeutics, offer children and adults suffering from these debilitating conditions the hope that breakthrough, potentially curative therapies may be a realistic goal in the future.

Three of the articles concern mood disorders, leading causes of morbidity and mortality (from suicide) worldwide. Major depression is very common, with a lifetime incidence of approximately 16% and a 12-month prevalence of 7%. In their review of novel therapeutic approaches for major depression, Sen and Sanacora note that the therapeutic properties of the currently available antidepressant medications, which include monoamine oxidase inhibitors, tricyclic antidepressants, selective serotonin reuptake inhibitors (SSRIs), and selective serotonin-norepinephrine reuptake inhibitors, were identified initially through serendipity. For example, the first monoamine oxidase inhibitor was derived from an antituberculosis medication in the 1950s, and it ushered in a psychopharmacological era primarily focused on drugs modulating the neurotransmitters norepinephrine, serotonin, and dopamine. However, despite more favorable tolerability, newer medications (such as SSRIs) have not provided substantial advances in efficacy or speed of onset and have not offered a rational pathophysiological basis for disease management. It is noteworthy that recent large-scale clinical trials have suggested that approximately 50% of patients do not achieve clinical remission following 2 treatment trials.[2]

Both this article and Zarate's review of bipolar disorder suggest that drugs from other medical specialties will continue to play an important role in the development of novel treatment paradigms for serious mood disorders. Although some have commented on the loss of serendipity in psychopharmacology,[3] the emergence of several nonmonoaminergic therapeutic targets for mood disorders may portend a new era of rational serendipity, in which drugs from various specialties of biomedicine are tested for their hypothesized antidepressant and mood-stabilizing properties. Compounds under investigation for mood

Address Correspondence to:

Sanjay Mathew
Mount Sinai School of Medicine
New York, NY
Email: sanjay.mathew@mssm.edu

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20049

© 2008 Mount Sinai School of Medicine

disorders, based on pathophysiological models of illness or hypothesized mechanisms derived from existing compounds,[4] include drugs used in neurology, gynecology, and oncology, such as riluzole, *N*-methyl-D-aspartic acid receptor antagonists, tumor necrosis factor alpha antagonists, mifepristone, and tamoxifen. Glutamate-modulating agents in particular may offer a promising approach for the rapid treatment of refractory mood symptoms.[5]

Although the preponderance of research in experimental therapeutics has been directed toward neurochemical abnormalities, until recently there has been relatively less focus on the neural circuits underlying these neurochemical abnormalities. On the basis of neuroanatomical hypotheses of mood regulation and response to antidepressant interventions, Mayberg and colleagues[6] selected the bilateral subgenual cingulate (Cg25) as the target brain region for deep brain stimulation (DBS) in a small group of patients with treatment-resistant depression (TRD). DBS is a nonablative neurosurgical procedure that has been used successfully to manage motor symptoms in advanced Parkinson's disease, essential tremor, and primary dystonia. The articles by Dumutriu and colleagues and by Berlin and Hollander in this issue discuss the rationale and preliminary data for DBS for TRD and obsessive-compulsive disorder, respectively. In 2008, 2 industry-supported, multicenter controlled clinical trials of DBS for TRD will be conducted in the United States, with one study targeting Cg25 and the other study targeting a region adjacent to the ventral capsule/ventral striatum. Although the initial reports for TRD are promising, with long-term efficacy reported in obsessive-compulsive disorder,[7] the field awaits the results of these large trials.

Hoffman and Mathew review the state of the science for the psychopharmacology of anxiety disorders. Despite the prevalence and morbidity of posttraumatic stress disorder (PTSD), there are only 2 Food and Drug Administration–approved medications for it, both of which have modest effect sizes (with particularly limited efficacy in subgroups such as combat veterans). Identifying new treatments beyond SSRIs for PTSD is undoubtedly a critical imperative. Stress-related conditions such as PTSD are associated with chronically increased activity of brain circuits that use corticotropin-releasing factor (CRF), a neuropeptide involved in mediating the neuroendocrine, immune, autonomic, and behavioral responses to stress. The development of CRF_1 receptor antagonists has thus been a focus of intense investigation for mood and anxiety disorders. A recent placebo-controlled trial of a CRF_1 antagonist in major depression was negative.[8] However, several early-phase studies of CRF_1 antagonists are ongoing and aim to provide the first evidence that these robust neurobiological findings have translational implications for patients with anxiety disorders.

Finally, perhaps no other topic in psychiatry is as controversial as the use of psychotropic medications in children and adolescents. In 2004, the Food and Drug Administration issued a black box warning for suicidality in the pediatric use of antidepressants, and this warning was extended last year to include young adults. Cardiovascular risks associated with psychostimulants, a drug class commonly used for attention deficit hyperactivity disorder, have prompted appeals to clinicians to curtail their administration.[9] These developments parallel the growing recognition that serious mental illnesses such as schizophrenia and bipolar disorder often manifest initially in childhood. The timely article by Ivanov and Charney reviews the efficacy and safety of antipsychotic medications in youth for schizophrenia and bipolar disorder and outlines some of the key risk-benefit considerations when these agents are being prescribed.

DISCLOSURES

Dr. Mathew has received lecture or consulting fees from AstraZeneca, Jazz Pharmaceuticals, and has received research support from Alexza Pharmaceuticals, NARSAD, NIMH, and Novartis.

Dr Mathew has been named as an inventor on a use-patent of ketamine for the treatment of depression. If ketamine were shown to be effective in the treatment of depression and received approval from the Food and Drug Administration (FDA) for this indication, Dr. Mathew could benefit financially.

Dr. Mathew is supported by NIMH grant K23-MH-069656.

REFERENCES

1. Druss BG, Hwang I, Petukhova M, et al. Impairment in role functioning in mental and chronic medical disorders in the United States. *Mol Psychiatry* 2008. DOI: 10.1038/mp.2008.13. Available at: http://www.nature.com.
2. Rush AJ. Limitations in efficacy of antidepressant monotherapy. *J Clin Psychiatry* 2007; 68(suppl 10): 8–10.
3. Klein DF. The loss of serendipity in psychopharmacology. *JAMA* 2008; 299: 1063–1065.

4. Mathew SJ, Manji HK, Charney DS. Novel drugs and therapeutic targets for severe mood disorders. *Neuropsychopharmacology* 2008. DOI: 10.1038/sj.np.1301652. Available at: http://www.nature.com.
5. Sanacora G, Zarate CA, Krystal JH, Manji HK. Targeting the glutamatergic system to develop novel, improved therapeutics for mood disorders. *Nat Rev Drug Discov* 2008; 7: 426–437.
6. Mayberg HS, Lozano AM, Voon V, et al. Deep brain stimulation for treatment-resistant depression. *Neuron* 2005; 45: 651–660.
7. Greenberg BD, Malone DA, Friehs GM, et al. Three-year outcomes in deep brain stimulation for highly resistant obsessive-compulsive disorder. *Neuropsychopharmacology* 2006; 31: 2384–2393.
8. Binneman B, Feltner D, Kolluri S, et al. A 6-week randomized, placebo-controlled trial of CP-316,311 (a selective CRH_1 antagonist) in the treatment of major depression. *Am J Psychiatry* 2008; 165: 617–620.
9. Nissen SE. ADHD drugs and cardiovascular risk. *N Engl J Med* 2006; 354: 1445–1448.

Experimental Therapeutics for Refractory Obsessive-Compulsive Disorder: Translational Approaches and New Somatic Developments

Heather A. Berlin, PhD, MPH,[1] Holly Hamilton, BA[1] and Eric Hollander, MD[1,2]

[1]Department of Psychiatry, Mount Sinai School of Medicine, New York, NY
[2]Seaver and NY Autism Center of Excellence, Mount Sinai School of Medicine, New York, NY

ABSTRACT

Significant advances over the past 20 years in our understanding of the phenomenology and pathophysiology of obsessive-compulsive disorder, made in part from structural and functional neuroimaging and genetics research, can guide treatments that target brain regions, circuits, and neurotransmitter systems specific to obsessive-compulsive disorder, the disruption of which may alleviate obsessive-compulsive disorder symptoms. We discuss here our current understanding of the underlying neurobiology and heritability of obsessive-compulsive disorder and integrate that understanding with a review of the current pharmacological, neurosurgical, and brain stimulation treatments of refractory obsessive-compulsive disorder. Expanding on these studies, we hope that new pharmacological and psychological treatment strategies and research-driven targets for lesioning, stimulation, or other types of focal neuromodulation can be identified that could lead to future research directions. Cross-species translational research and neuroimaging of the physiological and anatomical pathways implicated in the pathophysiology and treatment response in obsessive-compulsive disorder will advance our understanding of the neural basis of obsessive-compulsive disorder and lead to more targeted and effective treatment options. ***Mt Sinai J Med 75:174–203, 2008.*** © *2008 Mount Sinai School of Medicine*

Address Correspondence to:

Heather A. Berlin
Department of Psychiatry
Mount Sinai School of Medicine
New York, NY
Email: heather.berlin@mssm.edu

Key Words: caudate nucleus, cingulate gyrus, deep brain stimulation, internal capsule, neurosurgery, obsessive-compulsive disorder, prefrontal cortex, serotonin reuptake inhibitors, therapeutics, transcranial magnetic stimulation.

Obsessive-compulsive disorder (OCD) is a relatively common, chronic illness associated with considerable morbidity and economic and social burden.[1–5] OCD is characterized by intense anxiety caused by unwanted, intrusive, persistent thoughts, images, or impulses (obsessions), which lead to repetitive behaviors or mental acts (compulsions) that the person feels driven to perform to prevent or reduce his or her distress or anxiety.[6] Obsessions and compulsions are time-consuming and cause significant functional impairment and/or distress.[7] OCD has a mean lifetime prevalence of approximately 2% to 3% in the general population,[5,8–11] more than twice that of schizophrenia,[6] and is the fourth most common mental disorder after depression, alcohol and substance misuse, and social phobia.[12] Lifetime and annual prevalence rates of OCD are remarkably similar across cultures.[6,9]

OCD onset may begin in childhood, adolescence, or early adulthood.[13] The mean age of onset is in late adolescence for men and in the early twenties for women, for whom the incidence is slightly higher.[6,9] Onset is gradual for most, but acute onset has been noted in some cases. OCD may follow an acute, episodic, or chronic course.[14] Most people have a chronic waxing and waning course, with exacerbation of symptoms that may be stress-related. OCD is highly comorbid with a variety of disorders,[5] especially anxiety disorders,[14–18] depression,[19–21] alcohol or substance misuse,[22–24] eating disorders,[25] and body dysmorphic disorder.[26]

This article contains supplementary materials available at http://www3.interscience.wiley.com/journal/112736813/issueyear?year=2008

Published online in Wiley InterScience (www.interscience.wiley.com).

DOI:10.1002/msj.20045

© 2008 Mount Sinai School of Medicine

The most common treatments for OCD are pharmacological and cognitive behavioral interventions. According to the American Psychiatric Association treatment practice guidelines for OCD,[27] selective serotonin reuptake inhibitors (SSRIs) are considered first-line treatments for OCD. However, SSRIs are often associated with delayed onset of therapeutic effect (8–12 weeks), only partial symptom reduction, and response failure or intolerability in 40% to 60% of patients. Pharmacological options for SSRI refractory cases include increasing drug dose, changing to another SSRI or clomipramine, combining SSRIs, or changing the mode of drug delivery. Augmentation with second-generation antipsychotics has demonstrated efficacy as a second-line treatment.

Despite this, some patients still remain refractory to all standard pharmacological and psychological treatments. This review primarily focuses on several alternative medical interventions that have been considered for these severe cases, such as ablative neurosurgery, and brain stimulation techniques, such as electroconvulsive therapy (ECT), transcranial magnetic stimulation (TMS), and the nonablative neurosurgical procedure deep brain stimulation (DBS). The study of translational approaches, including underlying endophenotypes (mediating factors based on the study of neurochemistry, neurobiology, and cognition) and predictive animal models used to investigate genetic factors and drugs with anticompulsivity effects, is crucial to the development of appropriate psychological and somatic treatment methods for refractory OCD and to our understanding of the pathogenesis of the disorder.

To review the relevant original research studies, we conducted a computerized literature search of recent articles (PubMed, 2000 to February 2008, any language) and also included older articles (before 2000) that we deemed relevant. Various combinations of keywords were used. We also considered sources cited in the reports identified by our original search and some key reviews. Given the limited number of available studies in certain areas, particularly DBS, we applied no methodological exclusion criteria. However, given the breadth of these topics and in order to keep the article concise and focused, we report only those publications that we consider most pertinent to and as a framework for the alternative somatic treatments and those publications that were considered by the authors to meet rigorous research standards.

After briefly describing the proposed etiology of OCD in terms of the underlying neurobiology, genetics, and environmental and psychological factors, we go on to discuss animal models that seem most relevant to the development of new treatment approaches for OCD. We then describe standard first-line treatments for OCD and the treatments used for patients who are incomplete responders or nonresponders to serotonin reuptake inhibitors (SRIs). Finally, we discuss treatments for patients who are still treatment-resistant, namely, ECT, TMS, ablative neurosurgical procedures, and DBS, and conclude with a discussion of possible directions for future research.

ETIOLOGY

Neurobiology

Although the cause of OCD is unknown, there is increasing evidence for the involvement of biological factors.[4,5] For any individual, a range of factors are likely to contribute to the expression of the disorder. OCD is heterogeneous in terms of the types of obsessions and compulsions, heritability, and comorbid conditions, and this probably reflects heterogeneity in the underlying pathology.[28] In accordance, there are many disorders known as obsessive-compulsive spectrum disorders that share features with OCD, such as trichotillomania, Tourette's syndrome (TS), and body dysmorphic disorder.[26]

Although the neurobiological basis of OCD (symptoms and related cognitive impairments) is unclear, lesion, functional neuroimaging, and neuropsychological studies have implied that structural and functional dysfunction of limbic or affective corticostriatothalamocortical circuitry, which includes the orbitofrontal cortex (OFC), plays a key role.[28–31] These circuits, first identified in nonhuman primates,[32,33] have been also been identified in human lesion and imaging studies of OCD patients.[34–36] Reduced gray matter volume of OFC in OCD patients compared to healthy controls has been the most consistent finding from structural magnetic resonance imaging (MRI) studies. The OFC is involved in behavioral adaptation to change and motivational aspects of decision making,[37] both of which have implications for OCD. Similarities in neurocognitive deficits of OFC lesion[38] and OCD patients in a range of cognitive and behavioral tasks involving executive control further implicate the OFC in the pathophysiology of OCD.[39] Although studies suggest that OCD patients have increased activity in frontostriatal circuitry, both hypoactivity and hyperactivity of the prefrontal cortex (PFC) may produce deficits in perceptual and cognitive flexibility, but for different reasons.[40] There is also imaging evidence, although less consistent, of volume changes in caudate nucleus, anterior cingulate cortex, and medial temporal lobe structures.[36,41–54] The inconsistencies in the structural imaging literature may be related to methodological issues (eg, medication effects and small sample sizes),[55] the heterogeneity of OCD,[56] or the comorbidity of the patient samples.[43]

Like structural neuroimaging, functional imaging studies [positron emission tomography (PET), functional MRI, and single-photon emission computed tomography] have implicated several brain areas, including the OFC, the head of the caudate nucleus, and the thalamus.[36,56–64] Many studies have shown elevated blood flow/activation in frontal subcortical circuits in OCD patients compared to healthy controls.[64–66] In particular, PET and functional MRI have shown increased glucose metabolism in the OFC, caudate nuclei, and anterior cingulate in OCD patients. Cortical and basal ganglia regions have been most strongly implicated,[67] but a recent meta-analysis found differences only in the orbital gyrus and the head of the caudate nucleus.[64] Functional imaging data thus far suggest hypermetabolism of frontal regions,[55] and consensus is growing regarding the metabolic states of the caudate and thalamic regions.[68] In general, these findings suggest that frontal and subcortical regions play key roles in OCD,[64] but it is unclear whether these hypermetabolic states are the cause or result of OCD. For a comprehensive review of structural and functional neuroimaging studies of OCD, see Friedlander and Desrocher.[69] The findings are presented in the context of 2 models: executive dysfunction, implicating the dorsolateral PFC, caudate nucleus, striatum, and thalamus [see Supplementary Table 1 at the *Mount Sinai Journal of Medicine* Web site (http://interscience.wiley.com)], and modulatory control, implicating the OFC, medial PFC, and cingulate gyrus [see Supplementary Table 2].

However, there are some inherent biases in imaging studies investigating regions of interest (ROIs). They are likely to generate hypotheses related to the regions that they are studying (eg, OFC-striatal circuitry in OCD) but may miss important contributions from brain regions not investigated to date. However, recently developed whole brain–based structural imaging techniques, such as voxel-based morphometry (VBM) and multivoxel analyses, can examine differences in gray matter throughout the brain without the need to prespecify ROIs.[70,71] These unbiased whole brain (and potentially whole genome) techniques may yield previously unexpected new findings by revealing gray matter differences in areas not considered previously and can be used to confirm the OFC-striatal hypothesis developed in ROI structural MRI studies. Three VBM studies on OCD have been published to date,[49,72,73] and consistent with findings from ROI studies, they provide some evidence for orbitofrontostriatal structural abnormalities. However, 2 of the 3 VBM studies report structural changes in parietal regions: the right supramarginal and angular gyri[73] and the left inferior parietal lobe.[49] Thus, parietal lobe structural abnormalities should perhaps be a target for further investigation in OCD.

Several neurotransmitter systems have been implicated in the pathogenesis of OCD. Many studies demonstrating the efficacy of SSRIs in the treatment of compulsive behavior have supported the role of serotonin in OCD. It has been suggested that OCD patients have excessive baseline activity of excitatory glutamatergic neurons of the OFC, and because serotonin inhibits these neurons, an increased release of serotonin in the OFC may therefore be associated with a decline in OCD symptoms.[74] Studies investigating serotonin receptor function have also supported its pivotal role, and it has been found that OCD symptoms are exacerbated after the administration of *meta*-chlorophenylpiperazine, a serotonin 5-hydroxytryptamine receptor 2C (5-HT2C) and 5-hydroxytryptamine receptor 1D (5-HT1D) agonist,[75,76] and sumatriptan, a 5-HT1D agonist.[77] Recently, Adams *et al.*[78] found increased 5-hydroxytryptamine receptor 2A receptor binding in the caudate nuclei of untreated OCD patients.

The dopamine system also appears to play a role in the pathology of OCD, and many studies have suggested that antipsychotic augmentation in SSRI nonresponders may be an appropriate treatment for refractory OCD.[79] In addition, a recent neuroimaging study found increased dopamine transporter density in the left caudate and left putamen in OCD patients versus controls.[80] Evidence from animal models also points to the role of dopamine in compulsive behaviors (eg, quinpirole-treated rats[81]; see the Animal Models section).

Finally, several lines of evidence suggest that dysfunction in glutamate neurotransmission may contribute to the pathophysiology of OCD, although the precise abnormality in transmission is unknown.[82] An antagonist of the *N*-methyl-D-aspartic acid glutamate receptor has been shown to worsen the repetitive behavior of transgenic D1CT-7 mice.[83] D1CT-7 mice are engineered transgenic mice that express a neuropotentiating protein (cholera toxin A1 subunit) in a cortical-limbic subset of dopamine D1-receptor expressing (D1+) neurons. Furthermore, preliminary studies have suggested that glutamate modulating drugs such as topirmate and riluzole may constitute an effective alternative pharmacological strategy for treating OCD[84–91]; however, further controlled trials are warranted.

Genetics

Twin and family studies have provided evidence that genetic factors are involved in the transmission and expression of OCD.[8,92–103] Family studies have shown that the risk of developing OCD is 3 to 12 times greater in first-degree relatives than in the general population.[92,98,104,105] Furthermore, concordance for OCD is greater in monozygotic twins (80%–87%) than in dizygotic twins (47%–50%).[93,95] In a meta-analysis of family and twin studies, a person with OCD was reported to be 4 times more likely to have another family member with OCD than a person without OCD.[94] Genetic and family studies have also shown that OCD appears to be related to tic disorders and TS.[98,106–109]

Although twin and family studies implicate a genetic component for OCD, the specific mechanism of inheritance and the genes involved are not known.[94] This may be due to the etiological heterogeneity of OCD. Studies in the 1980s suggested an autosomal dominant pattern of inheritance for the combined phenotypes of OCD, TS, and chronic tics.[110] More recently, have studies have implicated a major locus with a multifactorial genetic contribution.[111,112] There is some evidence for linkage on chromosome 9p24,[104,113] but association studies examining candidate genes in the 9p24 region have been mixed.[113,114] The most promising candidate gene in the region, shared by the 9p24 linkage findings and the reported 9p monosomy, is solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag) member 1 (*SLC1A1*), which codes for the glutamate transporter excitatory amino acid carrier 1 (excitatory amino acid transporter 3).[114]

Many other studies have shown mixed results in examining genetic loci mainly associated with serotonergic, dopaminergic, or glutamatergic pathways or immunological processes as functional candidate genes for OCD.[115–121] Genetic polymorphisms of serotonin transporters have been implicated in OCD.[122] This coincides with the effectiveness of SSRIs in treating OCD and with the presence of high concentrations of serotonin receptors within the ventrolateral caudate nucleus.[123] Genetic polymorphisms in the number of variable tandem repeats of dopamine receptor genes, especially the dopamine receptor D_4 gene, have also been associated with OCD.[122] This coincides with the efficacy of dopamine blockers for OCD treatment and of dopamine blockers added to SSRIs for some refractory OCD patients[5] and with the functional role of dopamine in basal ganglia structures such as the caudate thought to be involved in OCD. These studies provide further evidence for the involvement of corticostriatothalamic circuits in OCD by suggesting that OCD may be mediated by an inherited dysfunction in serotonin and dopamine systems.[112] Recent genetic association studies of OCD in humans have also implicated genes involved in glutamatergic neurotransmission.[115,124,125] This, in conjunction with the synapse-associated protein 90/postsynaptic density-95-associated protein 3 (SAPAP3) study in mice,[126] suggests that defects in excitatory synaptic transmission in the corticostriatal circuit may contribute to the pathogenesis of OCD. For a review of human genetic studies of OCD, see Hemmings and Stein.[127]

Environmental and Psychological Factors

There are currently no established environmental risk factors for OCD. Several studies have reported major life events in the period preceding the onset of OCD.[128,129] Some work has demonstrated that psychological trauma may play a role in some cases of OCD.[130,131] Additionally, pregnancy[132] and the puerperium period[133] have been identified as potential risk factors for OCD, and cases of postpartum-onset OCD have been reported.[134] A recent study found that out of 56 female OCD patients who had been pregnant in the past, 38.2% reported that their symptoms began or changed during pregnancy or within a month of giving birth.[135] Furthermore, Williams and Koran[136] reported that in 7 of 24 women with preexisting OCD, symptoms were exacerbated in the postpartum period. However, rather than the events themselves being causal, a stressful life event may be a trigger in people who are biologically or psychologically predisposed to OCD. The role of stressful life events (including pregnancy and childbirth) as potential risk factors for OCD still needs further investigation.

Some have suggested that streptococcal infection is associated with a form of early-onset OCD.[137–144] Streptococcal infections trigger an immune response, which in some people generates antibodies that cross-react in the basal ganglia, a region implicated in OCD. However, the link between infections and exacerbation of symptoms is still questionable. For example, 1 study found no relationship between new streptococcal infections and symptom exacerbation in an unselected group of patients with OCD and/or TS.[145]

In terms of psychological models of OCD, some evidence suggests that the way in which people interpret their thoughts and some of their beliefs, including responsibility, the need to control thoughts, and thought-action fusion, may play a maintaining role in the disorder.[146–151] However, there is little evidence to date to suggest that these beliefs play a causal role in the etiology of OCD.

Animal Models

Animal models of OCD can be divided into 3 main classes: behavioral, genetic, and pharmacological. We have chosen to discuss here those models that seem most relevant to the development of new treatment approaches for OCD.[152,153] For a comprehensive review see Joel (2006).[152]

Behavioral Models

Behavioral models were the earliest animal models of OCD and include any naturally occurring stereotypic or repetitive behaviors (ie, tail chasing, fur chewing, and weaving); innate motor behaviors occurring during conflict, stress, or frustration (ie, grooming, cleaning, or pecking); and learned behaviors that become compulsive (ie, after behavioral manipulation such as signal attenuation–induced compulsive lever pressing).

In terms of naturally occurring stereotypic or repetitive behaviors, Powell *et al.*[154] showed that deer mice engage in stereotypic behaviors such as patterned running, backward somersaulting, and vertical jumping. In a study by Presti and Lewis,[155] high-stereotypy mice (in comparison with low-stereotypy mice) exhibited significantly decreased enkephalin content and significantly increased dynorphin/enkephalin ratios. The results are consistent with the hypothesis that spontaneous stereotypic behavior is a result of relative hyperactivity along cortico–basal ganglia–cortical feedback circuits involving the direct (facilitative) pathway, but they also suggest that this imbalanced activity involves disruptions of the indirect (inhibitory) pathway. This type of imbalance has been implicated in the compulsions of OCD.

Examples of innate motor behaviors occurring after behavioral manipulation include schedule-induced polydipsia, food restriction–induced hyperactivity, and marble burying. In 1 study,[156] polydipsia was induced in food-deprived rats by exposure to a fixed-time feeding schedule. Interestingly, chronic administration of SSRIs commonly used to treat OCD (ie, fluoxetine, clomipramine, and fluvoxamine) improved schedule-induced polydipsia in rats, whereas other drugs that do not improve compulsive behaviors (ie, haloperidol and diazepam) did not. Altemus *et al.*[157] treated rats with fluoxetine, imipramine (with no known effect on OCD symptoms), or saline prior to food restriction and exposure to a running wheel. Rats fed for 90 minutes per day could stabilize their weight after an initial decrease; however, if they were given a running wheel, they ran excessively, ate less, and lost weight. Rats receiving fluoxetine lost less weight, ran more, and increased food intake more rapidly than those who received saline. There were no differences in rats receiving imipramine and controls.

More common than schedule-induced polydipsia and food restriction–induced hyperactivity, is marble burying in mice. Rodents commonly use bedding material to bury objects that are both harmful and innocuous, and the discovery that marble burying is reduced by SRIs suggested that the behavior may be related to OCD and compulsive behaviors.[158–162] Londei *et al.*[161] suggested that this activity may begin as investigative, and because the marbles are nonreactive, they are unable to provide the necessary stop signal, and this "frustrated" investigation leads to compulsive burying. Similarly, compulsive behaviors may result from an inability to achieve a feeling of task completion.[163] It has been found, however, that burying is also reduced by drugs that do not affect compulsivity, such as diazepam,[158,159] and this had led to a lack of predictive validity that would accompany a selective response to SSRIs.

The signal attenuation model exemplifies learned behaviors that become compulsive-like after behavioral manipulation and stems from the suggestion that compulsive behaviors result from a failure to stop responding after the successful completion of an action due to inadequate response feedback mechanisms.[164] This may be induced with a post-training signal attenuation paradigm that leads to a pattern of compulsive-like lever pressing. Acute administration of SSRIs (ie, paroxetine and fluvoxamine) but not of drugs that do not improve OCD (ie, desipramine, diazepam, and haloperidol) has been found to decrease compulsive behaviors within the model.[165,166] Additionally, there is evidence of involvement of the serotonergic and dopaminergic systems in compulsive lever pressing, and strain differences in levels of compulsive lever pressing and resistance to the anticompulsive effect of paroxetine have been thought to parallel strain differences in dopamine and serotonergic function.[167] This, along with evidence of OFC involvement, falls in line with the neural systems implicated in OCD.

Genetic Models

Currently, 5 common animal models of OCD (described next) involving observed behaviors of genetically modified mice reflect compulsive and repetitive behaviors that are comparable to symptoms of OCD in humans [see Supplementary Table 3]. Unfortunately, these models are not supported by pharmacological treatment reports, which could prove invaluable in elucidating the neural systems involved in OCD.[152]

Campbell *et al.*[168] generated transgenic mice that expressed a neuropotentiating enzyme (an intracellular form of cholera toxin) within a cortical-limbic subset of dopamine D1 receptor–expressing (D1+) neurons. These mice, named D1CT-7 mice, exhibited compulsive behavioral abnormalities such as perseverance or repetition of normal behaviors, repetitive leaping, and nonaggressive repeated biting of siblings during grooming. The authors emphasized the similarities between the brain regions in which transgene expression was evident in the D1CT-7 mice and the neural substrates of human compulsive behavior (ie, the amygdala and limbic areas of the cortex). Another study found that D1CT-7 mice exhibit behaviors similar to TS, including juvenile-onset tics and tic responsiveness to clonidine, a drug used in humans for comorbid OCD and TS.[169] These results suggest that chronic potentiation of cortical and limbic D1+ neurons may cause compulsive behaviors similar to those in human compulsive disorders,[168] providing a transgenic mouse model of OCD and comorbid TS.

Mice with disruption of the homeobox B8 (*Hoxb8*) gene have been shown to exhibit excessive grooming behavior leading to hair removal and skin lesions, including frequent grooming initiation, increased duration of grooming, and grooming of normal cagemates.[170] This aberrant behavior in mutant mice is similar to the grooming common in trichotillomania and OCD, and interestingly, the *Hoxb8* gene is expressed in the OFC, anterior cingulate, striatum, and limbic system, all of which are implicated in OCD.

On the basis of previous observations that 5-HT2C receptor knockout (KO) mice not only become obese and hyperphagic but also frequently chew inedible objects,[171–173] Chou-Green *et al.*[174] examined 5-HT2C KO mice versus wild-type (WT) mice to look for additional compulsive behaviors, and they observed highly organized repetitive behaviors including frequent chewing (but not eating) of nonnutritive clay, "neat" chewing patterns on plastic mesh screens, and reduced habituation of head dipping in comparison with WT mice. The authors suggested that the organized manner of screen chewing represents an example of compulsive behavior similar to human checking, ordering, smoothing, or washing and that the slower habituation of head dipping may resemble compulsive checking in individuals with OCD. Thus, the authors concluded that the 5-HT2C receptor mutant mouse may provide a promising model of compulsive behavior in humans. Indeed, there is some evidence suggesting that 5-HT2C receptors may have an important role in OCD because 5-HT2C agonists and antagonists have been shown to make OCD symptoms worse,[76] although more studies of drugs useful in OCD would still be beneficial to further support the relationship between the behavioral abnormalities in these 5-HT2C KO mice and compulsive behaviors in humans.[152]

Berridge *et al.*[175] examined the grooming behavior of dopamine transporter knockdown (DAT-KD) mice (whose extracellular dopamine levels in the neostriatum rise to 170% of those in WT rodents) in order to examine whether a behavioral sequence in animals similar to OCD in humans shows sequential superstereotypy, which is defined as behavioral patterns of rigid sequential patterns of action involving dysfunction in the nigrostriatal dopamine systems. Monitoring grooming duration and the sequential pattern of the syntactic grooming chain (a fixed-action pattern consisting of 25 grooming movements serially linked in 4 phases and following 1 syntactic rule), the authors found that DAT-KD mice spent more time grooming than WT mice and that the serial pattern of the syntactic grooming chain in DAT-KD mice was strengthened and more resistant to interruption. Because of these observations, the authors suggested several similarities between DAT-KD mutant mice and OCD and TS patients. In sum, DAT-KD mice show aberrant behavior that may be related to disorders involving basal ganglia and dopamine dysfunction such as OCD and TS.

Recently, Welch *et al.*[126] showed that mice with genetic deletion of SAPAP3, a postsynaptic scaffolding protein at excitatory synapses that is expressed in the striatum (see Figure 1), exhibit increased anxiety and compulsive grooming behavior resulting in skin lesions and facial hair loss. Through electrophysiological, structural, and biochemical studies of these mutant mice, the authors revealed deficits in corticostriatal synapses, demonstrating the critical role for SAPAP3 at corticostriatal synapses and highlighting the importance of the corticostriatal system in compulsive behaviors like those of OCD. These behaviors were also assuaged by treatment with an SSRI, and this strengthens this model's relevance to OCD.

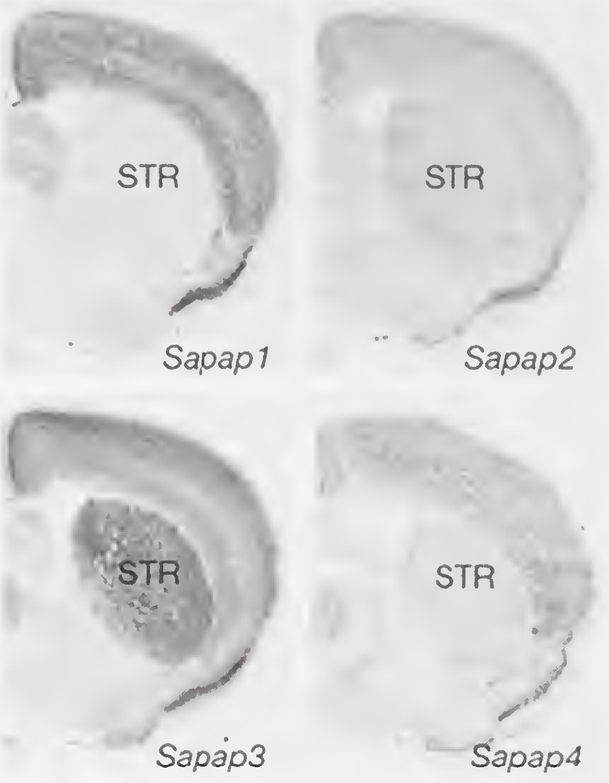


Fig 1. Only synapse-associated protein 90/postsynaptic density-95-associated protein 3 (Sapap3) messenger RNA is highly expressed in the striatum (STR), as shown here in SAPAP mutant mice. Reprinted by permission from MacMillan Publishers Ltd: Nature,[126] copyright 2007. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Pharmacological Models

Pharmacological models of OCD are based on drug-induced behavioral changes similar to symptoms of OCD, such as perseveration and indecision induced by manipulations of the serotonin system[176] and compulsive checking induced by changes in the dopamine system.[81,177,178] The quinpirole-induced compulsive checking model of OCD currently has the strongest face and construct validity.[152] In 1 study,[81] when placed in a large open field with 4 small objects in fixed locations, rats treated chronically with the dopamine agonist quinpirole revisited 2 locations excessively often and rapidly in comparison with those given saline. Furthermore, rats treated with quinpirole have been shown to perform ritual-like behaviors at these object locations.[81,178-180] The authors suggested that the behavior of quinpirole-treated rats is similar to compulsive checking in OCD patients because the rats met 5 ethological criteria for compulsive checking that could be suspended for a period of time (as compulsions may be in OCD patients) and because clomipramine was found to partially reduce the behavior.[81] Because dopamine abnormalities have been implicated in some obsessive-compulsive spectrum disorders and behaviors, this model is especially salient on account of its manipulation of the dopaminergic system to induce compulsive checking behaviors.

Summary of Animal Models

A variety of animal models have been proposed for compulsive behaviors and OCD, and each has its own strengths and shortcomings. Because many OCD patients do not respond to SRI therapy alone, it is necessary for a model to demonstrate a lack of effect on compulsive behavior by drugs typically found to not have an effect on symptoms of OCD.[152] The signal attenuation model does have this ability to discriminate between effects of effective drugs and those not known to improve OCD symptoms. Many of the other models discussed here, however, await further pharmacological research to enhance their predictive validity. The signal attenuation model also has strong construct validity stemming from the similarities in its mechanism and in the neural systems involved in OCD psychopathology. The genetic models help identify candidate genes and the neural systems underlying OCD and have strong potential in the realm of anticompulsive drug development, but these too will lack predictive validity until further research is conducted, and the question whether the behavior directly results from the manipulated gene or other targets downstream remains.[148] Combining results of various models may be helpful in elucidating the etiology and neurobiology of OCD and in developing novel treatment strategies.

TRADITIONAL TREATMENTS

The most common treatment approaches for OCD are pharmacological and psychological. Psychoanalytic psychotherapy and psychodynamic psychotherapy were the only psychological treatments for OCD for many years, but there is very little controlled evidence supporting their use. Cognitive behavioral therapy (CBT) was the first psychological treatment for which empirical support was obtained. CBT for OCD can be divided into (1) CBT that relies mainly on behavioral techniques, such as exposure and response prevention (ERP),[181,182] and (2) CBT that relies mainly on cognitive therapy, such as identifying, challenging, and modifying faulty beliefs.[183,184] However, these 2 forms of CBT as well as other cognitive and behavioral treatments generally overlap or are combined in clinical practice and treatment trials. A recent review compared psychological treatments with treatment as usual and found that psychological treatments derived from cognitive behavioral models are effective for adults with OCD.[185]

Most pharmacological treatments are aimed at regulating serotonin transmission. Serotonergic antidepressants such as clomipramine and SSRIs have been shown to effectively improve OCD symptoms.[186–188] In fact, clomipramine, fluoxetine, fluvoxamine, paroxetine, and sertraline, which are approved by the US Food and Drug Administration (FDA) for treatment of OCD, are recommended as first-line pharmacological agents.[27] Because SSRIs are more generally tolerable than clomipramine, an SSRI is usually preferred for a first medication trial.[27,189] The specificity of effectiveness of SSRIs in the treatment of OCD suggests that serotonin is an important neurotransmitter involved in the etiology and/or maintenance of OCD. However, studies using an acute tryptophan depletion paradigm have suggested that the treatment of OCD may be less dependent on the synaptic availability of serotonin than the treatment of other conditions such as depression. Acute tryptophan depletion produces a temporary reduction in serotonin synthesis and release, which has been found to cause remitted depression patients who have responded to serotonergic agents to relapse.[190,191] OCD patients in remittance, both SSRI-treated[192,193] and unmedicated,[194–196] however, have been found to evade relapse during acute tryptophan depletion. Therefore, the maintenance of treatment effects on obsessive-compulsive symptoms may not rely solely on the short-term availability of serotonin. Still, SSRIs remain the first-line pharmacological treatment for OCD.

Only 7 randomized trials have directly explored whether the combination of an SRI and CBT consisting of ERP is superior to either treatment alone in adults with OCD, but methodological limitations prevent definitive conclusions.[197–203] However, it has been suggested that a combination treatment (SRI plus CBT consisting of ERP) is appropriate when there are co-occurring disorders that are SRI-responsive, when there has been a partial response to monotherapy,[204] or when one is trying to reduce the chance of relapse when medication is discontinued.[27,205] Some research has indicated that psychological treatments such as CBT are as effective as antidepressants in causing regional brain metabolic changes correlated with OCD symptom improvement.[206] Both pharmacological and CBT treatments have been associated with a reversal of the functional neuroimaging findings (in terms of the glucose metabolism rate) to the pattern found in healthy controls.[207] However, the exact mechanisms of this reversal are not well understood. Findings suggest changes in caudate nucleus function with pharmacological and behavior therapy for OCD. In 1 study, improvement in OCD treated with behavior modification or fluoxetine hydrochloride was accompanied by significant changes in glucose metabolic rates in the caudate nucleus, as measured with PET.[206] Local cerebral metabolic rates for glucose in the head of the right caudate nucleus decreased significantly in comparison with pretreatment values in responders and in comparison with nonresponders and normal controls whose right caudate nucleus metabolism did not change from baseline. In another study, behavior therapy (structured ERP behavioral and cognitive treatment) responders had significant bilateral decreases in caudate glucose metabolic rates that were greater than those in poor responders.[207] Furthermore, the pretreatment correlations of brain activity between the orbital gyri and both the head of the caudate nucleus and the right thalamus decreased significantly after effective treatment.

Treatments for Incomplete Responders or Nonresponders to SRIs

Although considered the first-line and most effective and well-established treatment for OCD,[189,208,209] SSRIs are associated with several limitations, including a delayed onset of therapeutic effect (8–12 weeks of treatment to baseline) in most patients, only partial symptom reduction [25%–35% reduction in the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) from baseline] in some,[210,211] and failure to respond or tolerate SSRI treatment (after a single or multiple trials) in 40% to 60% of patients.[211–217] Y-BOCS[218,219] is an established obsessive-compulsive scale that measures the severity and change in severity of OCD symptoms. Specific criteria used to diagnose treatment resistance have been published elsewhere.[212,220,221] SRI-unresponsive OCD patients have substantially impaired social and occupational functioning,[212,222,223] and treatment-resistant OCD represents an ongoing challenge for researchers and clinicians. Over the last decade, different forms of treatment have been investigated, including pharmacological augmentation (eg, serotonergic and dopaminergic), psychotherapeutic augmentation, and alternative somatic strategies such as DBS (see the DBS section).

Pharmacological options for SRI refractory cases include increasing the drug dose, changing to another SSRI[189] or clomipramine, combining SRIs, and changing the mode of drug delivery (eg, intravenous versus oral clomipramine; for a review, see Fineberg and Gale[224]). However, augmentation with second-generation antipsychotics (dopamine antagonists) appears to be a promising strategy for SRI-resistant OCD. Open and placebo-controlled studies of antipsychotic augmentation of SRI monotherapy have shown very good results[225–228] in comparison with augmentation with serotonin-enhancing agents such as lithium, clonazepam, and buspirone.[229–233]

Randomized, placebo-controlled augmentation trials of both first-generation (haloperidol[231,232]) and second-generation (risperidone,[233–235] olanzapine,[236,237] and quetiapine[238–241]) antipsychotics have yielded response rates of 40% to 55% within 4 to 6 weeks.[233–240] Some controlled trials did not find significant differences between antipsychotic and placebo augmentation,[237,239,240] but methodological limitations may have contributed to the negative findings. In a recent review of antipsychotic augmentation in double-blind, randomized, controlled clinical trials, one-third of treatment-refractory OCD patients showed a meaningful treatment response to antipsychotic augmentation, and patients with comorbid tics were even more likely respond.[242] Still, it remains unclear how long patients need to remain on augmented treatment. A retrospective chart review[243] found that 15 of 18 patients (83%) who responded to antipsychotic augmentation relapsed within 1 year after the antipsychotic was discontinued, and 13 of the 15 who relapsed did so by the eighth week after discontinuation.

Some anticonvulsants (eg, valproate, oxcarbazapine, carbamazepine, gabapentin, and topiramate) have been reported to help patients either as monotherapy or as augmentation.[244–247] For example, in the first double-blind, placebo-controlled trial of topiramate augmentation of SSRI for the treatment of OCD, our group found that, compared to the placebo group (n = 18), the treatment group (n = 18) exhibited a significantly greater decrease in Y-BOCS compulsions over the 12-week study period with topiramate augmentation.[91]

Alternative Somatic Treatments

A minority of people with OCD who remain refractory to all standard pharmacological and psychological treatment (but not exclusively) are considered candidates for several alternative medical interventions such as ECT, TMS, ablative neurosurgical procedures, and nonablative neurosurgical procedures, namely DBS. Vagus nerve stimulation, used for conditions such as depression,[248] could also be a potential treatment for OCD, but because it has not been investigated for the treatment of OCD to date, it is not discussed here. The criteria used to establish treatment resistance and to determine eligibility for surgical treatments (stereotactic lesion procedures and DBS) in OCD patients have been published elsewhere[212,220,221,249] but include previous and often repeated failed trials of pharmacotherapy and psychological therapies.

Data for the treatment of OCD with these alternative somatic therapies are limited, as there is an understandable scarcity of double-blind trials. It is difficult to perform randomized control trials of invasive procedures with credible sham procedures, so most of the available reports are case series and open-label trials. Consequently, most studies reviewed here were based on a small group of a select sample of patients and were conducted at specific sites, often by a small group of people.[250] Although there are more data on stereotactic lesion procedures[251] than the other somatic therapies in treatment-resistant/intractable OCD, the cost, irreversibility, and lack of a clear relationship between specific anatomic lesions and successful outcomes limit their use.

In the past decade, there has been increasing interest in brain stimulation techniques in treatment-resistant neuropsychiatric conditions such as OCD and major depression. Stimulating focal brain regions (cortical and subcortical) either directly or indirectly with electrical currents may affect higher cognitive processes and mood systems, and this could explain its efficacy.[252] The most widely used brain stimulation technique is ECT, which was introduced in psychiatry over 70 years ago. ECT is one of the most effective treatments for severe, treatment-resistant depression, but its efficacy for the treatment of OCD is not as well established, and given the risks of the procedure and of general anesthesia, it is not clear if the risks outweigh the benefits. Newer methods of electrical brain stimulation, such as TMS and DBS, have a potential advantage over ECT in that they are able to stimulate brain regions more focally.

Mounting evidence from the genetics and imaging literature (discussed previously) suggests that several subcortical structures play a key role in OCD. In accordance, functional modification of these structures may reduce OCD symptoms. Therefore, advances in our understanding of the underlying neurobiology of OCD can guide treatments that target brain regions, circuits, and neurotransmitter systems specific to OCD, the interruption of which may alleviate OCD symptoms. Having discussed our current understanding of the neurocircuitry/neuroanatomy of OCD, we now integrate that understanding with a review of the targets used in ablative and brain stimulation treatments of refractory OCD. Based on the studies reviewed here, perhaps new research-driven targets for lesioning, stimulation, or other types of focal neuromodulation can be identified that could lead to future research directions.[112]

ELECTROCONVULSIVE THERAPY

ECT is a well-established treatment for severe depression,[253] but its efficacy in treatment-resistant OCD is still not established as published data are scarce, come mostly from case-report series, and show mixed results. The studies of ECT in treatment-resistant OCD include a retrospective case series,[254] 1 open trial,[255] and several individual cases[256–263] devoid of standardized measures, with some reported degree of effectiveness. However, the lack of standard outcome measures, absence of blinded trials, need for repeated anesthesia, and side effects of ECT preclude it from being a viable option for treatment-resistant OCD without comorbid conditions.[27,253] Furthermore, given the high rates of comorbidity in these published cases, ECT may reduce OCD symptoms by treating comorbid disorders such as depression,[264] TS,[265] and schizophrenia,[266] rather than affecting the OCD symptoms directly.

Therefore, given the methodological weaknesses (eg, high comorbidity and no blinded trials) and paucity of convincing evidence for sustained improvement in the ECT studies for treatment-resistant OCD thus far, in addition to the potential risks of ECT, a recommendation for the use of ECT in the treatment of pure treatment-resistant OCD cannot been made at this time.[267] However, ECT could be used to target comorbid conditions, such as depression, that might be exacerbating OCD symptoms. Alternatively, magnetic seizure therapy, a type of convulsive therapy in which the electrical stimulus used in ECT is replaced by a magnetic stimulus,[268,269] might offer a more focal stimulation with more benign side effects and thus calls for further investigation as it has not yet been tested in OCD patients.[270]

TRANSCRANIAL MAGNETIC STIMULATION

TMS, a noninvasive technique developed in 1985,[271] was first to used to treat neuropsychiatric conditions in 1987,[272] and since then, an increasing number of studies have investigated its efficacy in a range of neuropsychiatric illnesses.[273] TMS delivers magnetic pulses to the cortex via a stimulating coil (handheld or held by an external coil holder) applied directly to the head, which, depending on stimulation parameters [frequency (<1 to 20 Hz), rate, and duration], can enhance or decrease the excitability of the specific cortical region targeted[274–276] and modify regional cerebral blood flow.[277,278] In contrast to ECT, magnetic fields pass through the scalp and skull without the obstruction encountered by direct application of electricity, so less electricity is delivered to the brain in TMS;[279] this allows it to stimulate cortical regions more focally and with fewer side effects. Repetitive transcranial magnetic stimulation (rTMS), by which stimulation is delivered in trains of pulses (multiple stimuli per second applied to the same brain area for several consecutive seconds), is typically used in clinical practice. In OCD, TMS aims to modify PFC activity in order to influence obsessive-compulsive symptoms.

In sum, findings from the 4 published rTMS treatment trials for OCD to date [see Supplementary Table 4] are inconsistent, perhaps because of methodological differences between studies. In addition, the samples were small, and there was only 1 double-blind trial. Along these lines, a recent systematic review of TMS treatment for OCD concluded that there was not enough evidence from randomized controlled trials to determine its efficacy.[280] Thus, more research is needed to make any firm conclusions about the efficacy of TMS in OCD. Future studies should be better matched in terms of design, stimulation sites, treatment duration, and stimulation parameters. Although limited, the results thus far offer some promise, and the technique's noninvasiveness in comparison with more invasive techniques such as DBS and neurosurgery, its good tolerability (no dropouts and only temporary mild side effects), the lack of the need for anesthesia, and the relative ease of conducting double-blinded trials should encourage future research to better determine its efficacy.

Given the well-established involvement of subcortical circuits in the pathogenesis of OCD, doubts about the efficacy of TMS in OCD treatment have been raised because TMS penetrates only as deep as 2 cm; this makes the cortex its main target. However, deeper and more distant brain regions may still be affected by this cortical stimulation. In fact, studies combining TMS with functional neuroimaging have shown effects of TMS that are distant from the stimulation site.[277,281–283] In addition, coils with deeper penetration powers are being investigated for potential clinical use.[284] Doubts have also been raised by the absence of well-established localization for the stimulation. There are published reports of symptom reduction after stimulation in both the right and left PFC and the supplementary motor area, but a standardized target area for stimulation is still to be determined. Additional clinical trials could help identify the best position for the coil placement. In practical terms, the need for daily treatment may limit the use of TMS in clinical practice, and clinicians should be prepared for the very rare possibility of a seizure, the most severe adverse event occurring with rTMS, which may be more common at high stimulation frequencies. Hence, if better targets are established via new research and the technology of TMS advances to have a more focal effect, lower frequencies may be used, which may in turn reduce the risk of seizure.

Ablative Neurosurgical Procedures

Neurosurgical treatment for OCD is a highly selective treatment performed for relatively few patients with severe, treatment-refractory OCD when pharmacological and psychotherapeutic alternatives have been exhausted.[251] The expert consensus guideline recommends neurosurgery for treatment-refractory OCD as an "infrequently needed, but sometimes life saving intervention" when patients have been nonresponsive to 3 or more trials of SRIs (including clomipramine) and to CBT.[189] Of these patients who remain severely ill (~10–40% of OCD patients),[285–287] some are eligible for surgical intervention as long as they meet the appropriate inclusion criteria. The availability of reversible and adjustable DBS (discussed later) may lead to a decrease in the use of ablative neurosurgical procedures. However, these procedures still represent a potentially effective alternative for a select group of patients with very severe OCD.

Reviews of ablative neurosurgery for OCD have described promising results. In 1 study, 58% of 478 patients from 24 studies between 1961 and 1980 showed marked improvement, but more than half of the operations reviewed did not use stereotactic guidance.[288] In another review of 12 studies from 1961 to 1988, 67% of the 300 patients reviewed were categorized as "symptom free" or as having "minor symptoms," and 9 of the 12 studies reviewed used stereotactic guidance.[289] Finally, Freeman *et al.*[250] found an identical result; that is, 67% of 198 patients reviewed from the 5 studies fell into those categories.

Lesioning for the treatment of OCD is performed mostly via thermocoagulation (the use of heat produced by a high-frequency electric current to destroy tissue locally). The main anotomical targets for lesioning include the fiber tracts that connect the cortex to thalamic nuclei (subcaudate tractotomy), the anterior limb of the internal capsule (anterior capsulotomy; also performed via radiosurgery, which is known as gamma knife capsulotomy), and the cingulate gyrus (anterior cingulotomy). There is also a multitarget procedure in which lesions are made to areas that correspond to those made in both a cingulotomy and subcaudate tractotomy (limbic leukotomy; see Figures 2 and 3). See Supplementary Table 5 for a summary of the literature for the 4 main ablative procedures used in OCD.

In general, the quality of available evidence concerning the efficacy and safety of neurosurgical treatments is variable and inconclusive. There are a variety of design issues that prevent strong inferences from being made. For example, given the relative rarity of these interventions, studies are generally small (preventing identification of predictors of treatment response) and are conducted over long periods of time, and none have control conditions. However, it is unlikely that credible sham procedures would be considered ethical for ablative procedures, and patients are unlikely to accept randomization for these last-resort treatments. Evidence is also inconclusive on whether adverse

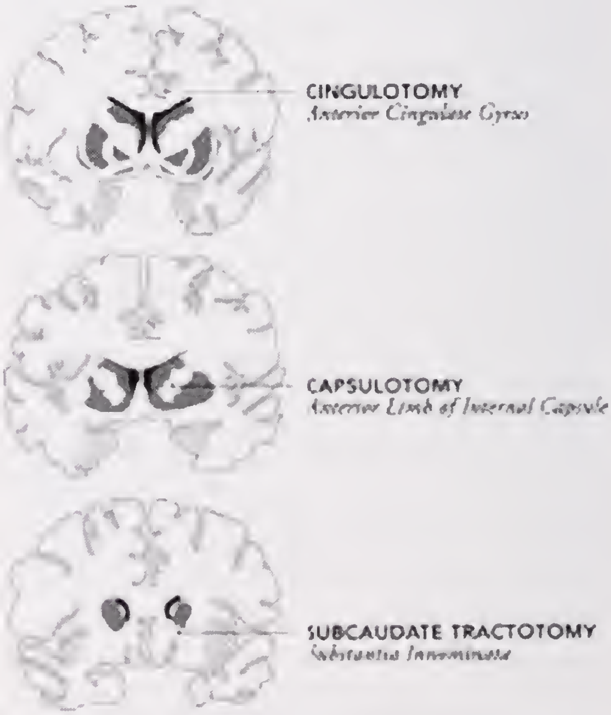


Fig 2. Anatomic locations of lesion targets thought to be important in the mediation of obsessive-compulsive disorder. Reprinted with permission from Lipsman *et al.* Deep brain stimulation for treatment refractory obsessive-compulsive disorder: The search for a valid target. Neurosurgery 2007; 61: 1–11.[112]

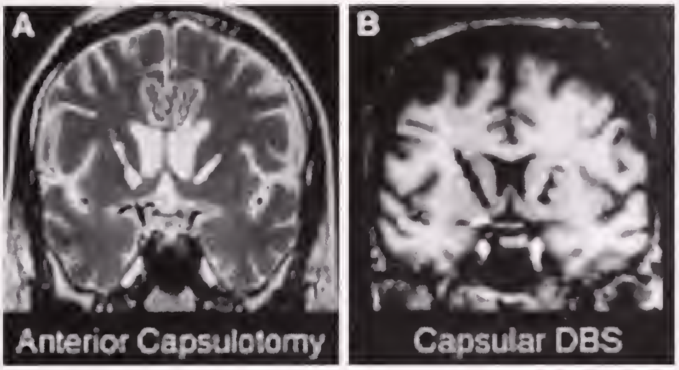


Fig 3. (A) Coronal T2-weighted image showing the lesion location in the anterior internal capsule for a capsulotomy using thermocoagulation. Reprinted from Neurosurgery clinics of North America, 14, 267–274, Copyright 2003.[314] with permission from Elsevier. (B) Coronal T1-weighted image showing the location of stimulating electrodes for deep brain stimulation (DBS) in the same location. Reprinted with permission from Lipsman *et al.* Deep brain stimulation for treatment refractory obsessive-compulsive disorder: The search for a valid target. Neurosurgery 2007; 61: 1–11.[112]

changes occur in neuropsychological and personality function in OCD patients who have received neurosurgery.

Furthermore, although there are reports of improvement for the 4 neurosurgical procedures described here, there are also reports of both transient and persistent adverse effects that should be taken into account. Even for studies with the strongest supportive clinical evidence, the proportion of patients who respond must be considered in the context of the potential for serious adverse events inherent in all neurosurgical procedures, including persistent ones as reported in some studies. The entry criteria in the more recent prospective series all include previous failure of pharmacological and psychological (eg, ERP) treatment trials, and this indicates that in general, at least in the last 15 years, these have been considered last-resort treatments.

As a result of contrasting results and varying inclusion criteria, there is also a lack of consensus in the literature on the most appropriate target for ablative surgery, and there is no convincing evidence to date comparing different neurosurgical procedures.[112] Several studies have suggested that multiple lesions may be needed for both cingulotomy and capsulotomy, which may increase the chance of response but may also increase the risk of adverse events. Although stereotactic lesion procedures have more empirical data[251] than the other somatic therapies (discussed later) in patients with treatment-resistant or intractable OCD, the cost, irreversibility, and lack of a clear relationship between specific anatomic lesions and successful outcomes limit their use. Still, for a small number of patients with the most severe, chronic, disabling forms of treatment-refractory OCD, most of whom also have significant psychiatric comorbidity, including severe depression, even the relatively modest degrees and rates of response reported in these studies may be clinically relevant. Therefore, these patients should continue to be assessed as candidates for neurosurgery, which may include nonablative procedures such as DBS (described next).

DEEP BRAIN STIMULATION

The potential long-term adverse effects and irreversibility of ablative neurosurgical procedures led to the investigation of less destructive neurophysiological interventions such as the relatively new technique of electrical DBS.[290] DBS is a nonablative neurosurgical procedure in which electrodes are implanted in the brain to stimulate brain regions directly. The stimulating leads are connected by an extension wire to pulse generators usually implanted under the skin of the chest, just under the collarbone (Figure 4). Stereotactic imaging with MRI and CT[291] are used for the precise anatomical localization of the electrodes. The effects of electrical brain stimulation were previously explored during stereotactic surgery for OCD before lesions were made (eg, see Fux *et al.*[339]). However, the therapeutic use differs in that electrodes are actually implanted in the brain, which are then stimulated by an external electrical source.

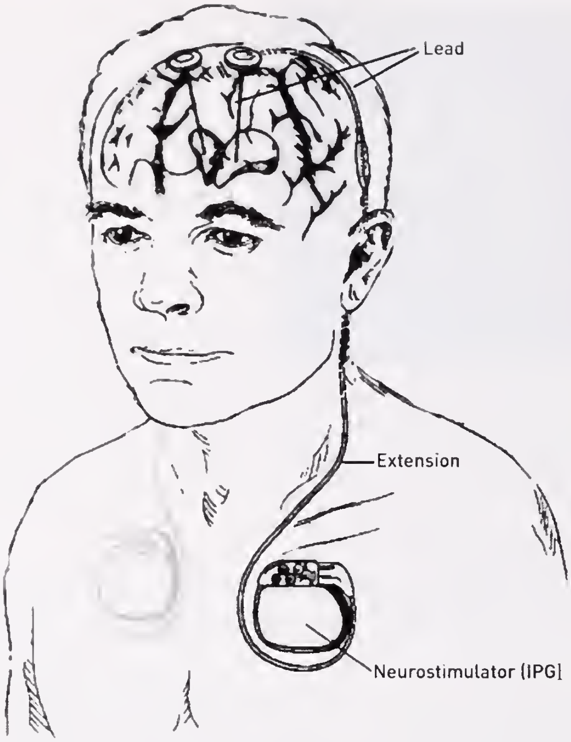


Fig 4. Illustration of the deep brain stimulation devices. Reprinted with permission from *CNS Spectrums*.[270] Copyright 2005, MBL Communications.

Any lesions that occur during the intervention or through repeated stimulation are not deliberate and are considered to be small with respect to those made deliberately during ablative procedures.

DBS has a number of advantages over traditional lesioning procedures. Because thermocoagulation destroys brain tissue in addition to offering a possible clinical benefit, it can produce harmful, irreversible consequences. In contrast, DBS has a much lower rate of side effects than lesioning with thermocoagulation,[292] and most importantly, DBS can produce a positive clinical effect without creating an irreversible lesion. DBS is reversible, in that the stimulation can be modified (adjusted up or down and delivered intermittently or continuously) at any point after the implantation or discontinued if side effects occur, and the device itself can be removed. The ability to modulate or turn stimulation on and off allows researchers to perform sham stimulations to measure placebo effects; this has not been possible for ethical and practical reasons with ablative procedures.[293] Therefore, DBS is a minimally invasive neurosurgical procedure that comes with the ability to control side effects from overstimulation and to completely remove the electrode without causing damage to the stimulated region. However, DBS still has risks, and even apparently simple aspects of the procedure can have harmful consequences if it is performed improperly.

The precise mechanisms underlying the efficacy of DBS are still not known (for a review, see Lozano *et al.*[294] or McIntyre *et al.*[295]), but they are most likely complex and involve more than just 1 mechanism to achieve the clinical effects observed.[295,296] By using high-frequency pulses (≥100 Hz), DBS is thought to have complex effects that include a blocking effect on the stimulated area, mimicking the effect of tissue lesioning[297,298] but without actually destroying tissue.[299] A number of mechanisms have been proposed, such as the release of inhibitory neurotransmitters, synaptic inhibition, depolarization blockade, synaptic fatigue/depression, neural jamming, and stimulation-induced modulation of pathologic network activity.[295,296,300] Because the effects of DBS are very similar to those of lesioning, the underlying mechanism probably involves an interruption of pathological neural activity, which varies, depending on the local circuitry and its neurotransmission, and may result from exciting or inhibiting axons and neurons.[296,301–304]

Even though its mechanisms are not yet fully understood, DBS has had very good benefits in patients with Parkinson disease (PD) and other movement disorders[305,306] in terms of reducing abnormal movements, and its efficacy and safety are well established for the treatment of these neurological conditions (see Hamani *et al.*[307]). Because it is reversible and adjustable, DBS is thought to be a safer, more conservative treatment than lesion surgery for these patients, especially those who need bilateral procedures.[308] In fact, DBS was recently approved by the FDA (in 1997 for unilateral implantation and in 2002 for bilateral implantation) for the treatment of intractable motor disorders (tremor[292,309,310] and PD[311]).

The relative success of DBS in the treatment of PD, its reversibility in comparison with ablative procedures, its relatively safe side-effect profile, and the suggested relationship between corticostriatothalamocortical loop dysfunction–mediated psychiatric and movement disorders have prompted researchers to investigate DBS as an alternative treatment for some psychiatric illnesses, including refractory OCD. DBS was developed as a treatment for OCD through collaboration between Belgian and Swedish researchers. The rationale for using DBS in this population is based on the empirical results of lesion studies and neuroimaging findings in OCD patients showing increased activity in neuronal circuits involving the medial PFC, OFC, anterior cingulate gyrus, and caudate nucleus (or, more generally, in the frontal–basal ganglia–thalamic circuit).[312] Because the blocking effect of high-frequency DBS on the stimulated area mimics the effect of lesioning tissue.[297,298] DBS applied to specific regions (eg, the anterior limbs of the internal capsule) theoretically interrupts these hyperactive circuits in the same way as traditional neurosurgical procedures, but without causing irreversible damage to brain tissue.

Studies of DBS for OCD

For a summary of DBS studies for the treatment of OCD, see Supplementary Table 6. In the first published report of DBS for treatment-resistant OCD,[313] 3 of 4 patients who were treated with chronic electrical stimulation instead of bilateral capsulotomy improved (little detail was provided). One of the 3 patients, who had severe OCD for over 20 years, reported an almost instantaneous relief of anxiety and obsessive thinking when stimulation was on, which disappeared when it was turned off. Quadripolar electrodes were stereotactically implanted bilaterally in the anterior limbs of the internal capsules (identical to the target in capsulotomy; Figure 3). DBS of the anterior capsule is thought to disrupt corticostriatothalamocortical loop fibers that connect the cortex and thalamus, which is thought to be a pathological circuit in OCD.[112]

Subsequently, a new series of 6 patients with severe treatment-refractory OCD treated with DBS has been reported in several publications.[314–316] Again, quadripolar electrodes were stereotactically implanted bilaterally in the anterior limbs of the internal capsules, but this time, 4 patients completed a double-blind crossover trial (ie, the stimulator was continuously on for 3 months and off for 3 months or vice versa in random order) and then had continuous stimulation thereafter. Three of the 4 patients responded (> 35% improvement on Y-BOCS and much improved Clinical Global Impression Improvement scores) during the stimulator-on condition versus the stimulator-off condition. Responders reported clinically meaningful improvement in the first week of stimulation, and their symptoms worsened during the course of a few days in the stimulation-off condition. There was a dose-response effect with stimulation, and the stimulation-induced effect (with continuous stimulation) was maintained for at least 21 months after surgery; this made a placebo effect unlikely (a study of PD patients found that the estimated magnitude of the placebo effect in DBS was equivalent to 39% of the magnitude of the effect of active DBS[317]). However, given the small sample size, these results should be interpreted cautiously. Side effects included transient disinhibtion with high amplitudes, fatigue, memory disturbances, and weight loss and gain, and 1 patient was aware of ("felt") the leads and the stimulator and was irritated by the thought of having implanted electrodes.[315] The main technical problem was the short battery life, which required replacement of the stimulators every 5 to 12 months.

A comparison of preoperative and postoperative PET scans of these patients showed a marked decrease in frontal metabolism after 3 months of stimulation. This coincides with functional imaging studies suggesting that OCD symptoms are mediated by hyperactive orbitofrontal-subcortical circuits, perhaps because of an imbalance between direct and indirect striatopallidal pathways.[67] Thus, bilateral DBS in the anterior limbs of the internal capsules may improve OCD symptoms via its effect on the activity of frontal-subcortical brain circuitry.

Two patients were not crossed over in the original study (one had a capsulotomy, and one was still in the postoperative screening phase) but completed the study subsequently.[315] Patient 5 received a different electrode placement: one in each internal capsule and one in each dorsomedial thalamic nucleus. The dorsomedial thalamic nucleus was not an effective target for this patient, who also did not respond to longer term internal capsule stimulation. When stimulation was turned on, patient 6 had a large improvement (> 50% decrease in postoperative tests) in his mood and aggressive, intrusive thoughts. However, his obsessions returned with their former intensity 1 week after the crossover to the stimulation-off condition, and he became very depressed and suicidal, so the blinded research team decided to end treatment.

Gabriels *et al.*[316] evaluated the impact of DBS on emotions, behavior, personality traits, and executive function in 3 patients from this series, 2 of whom experienced sustained improvement of OCD symptoms with DBS and a decrease in depression and tension with increasing stimulation amplitude. The Total Maladjustment Score on the Brief Psychiatric Rating Scale decreased by 44% and 59%, and this reduction in psychopathology was persistent under continuous stimulation. DBS did not have a negative effect on neuropsychological test performance or self-rated personality traits, and no harmful side effects were detected during follow-up (33–39 months). The lack of response in the third patient may have been related to his comorbid somatoform disorder and fixation on the thought of the electrodes in his head and the implanted stimulator under his skin. Upon removal of the DBS equipment and capsulotomy, his Y-BOCS scores and somatic complaints gradually decreased.

In contrast to the Belgian group, which selected the internal capsule as the target, Sturm *et al.*[318] chose the shell region of the right nucleus accumbens (NA) as the target for DBS in a pilot series of 4 patients with severe treatment-resistant OCD and anxiety disorders. Because the NA sits between the amygdaloid complex, basal ganglia, mediodorsal thalamic nucleus, and PFC, which are all involved in the pathophysiology of OCD[65] and anxiety disorders,[319] the clinical benefits of anterior capsulotomy may be due to blockage of amygdaloid–basal ganglia–PFC circuitry at the level of the shell of the NA rather than at the fiber tracts in the internal capsule.[318] Therefore, Sturm *et al.* implanted the electrodes so that the contacts could alternate between internal capsule and NA stimulation and both sites could be compared. For the first patient in this series, electrodes were implanted bilaterally, and activation of various contact combinations was alternated over several weeks. Stimulation of the right NA produced a significant reduction in symptoms, bilateral stimulation of the NA did not improve the effects, and stimulation of the internal capsule was not effective. Therefore, the electrode was implanted only unilaterally into the right NA in the subsequent 3 patients.

Three of 4 patients had a nearly complete recovery from both anxiety and OCD symptoms without any side effects with follow-up periods of 24 to 30 months. An MRI of the fourth nonresponsive patient showed that the electrode was displaced and missed the target area, and this may explain his lack of response. Although these results are promising, the authors did not state how the initial diagnosis and recovery were attained. Therefore, additional studies with more well-defined methodological information and a larger sample size are needed.

A PET study was conducted under stimulation-on versus stimulation-off conditions in 1 patient from this series.[318] High-frequency stimulation of the NA shell activated the right dorsolateral PFC and cingulate cortex but inhibited ipsilateral dorsolateral rostral putamen activity. As a central relay structure, the NA may help to modulate information flow from the amygdaloid complex to the amygdala, basal ganglia, mesolimbic dopaminergic areas, mediodorsal thalamus, and PFC via dopaminergic transmission.[320–322] On the basis of the overall clinical results of this series and the PET study, the authors suggested that OCD and anxiety disorders result from imbalanced information flow from the amygdaloid complex related to a disturbance in the NA, which modulates amygdalo–basal ganglia–PFC circuitry, and this can be rectified by the blocking of the information flow within the shell of the NA with DBS.

In a recent study from researchers at the University of Michigan,[323] bilateral DBS of the anterior limbs of the internal capsule was examined in 4 refractory OCD patients. This study consisted of a short-term, blinded, randomized, on-off design of four 3-week blocks and a long-term, open stimulation follow-up (up to 1 year). One patient exhibited marked improvements in mood, anxiety, and OCD symptoms (>35% decrease on Y-BOCS) during the blinded study and open, long-term follow-up. Another patient had moderate benefit during open follow-up. However, amplitude voltages used in the responding patients were on the high end, ranging from 5 to 10.5 V; for example, the patient with the greatest improvement had a setting of 7 V. Side effects were mostly transient and prominent only at high amplitudes and included tingling, nausea, and diarrhea. PET scans showed OFC deactivation in the 2 patients who responded. Thus, the authors suggested that the positive clinical response was mediated by a disruption of the cortical-subcortical circuit. It should be noted that one of the responders committed suicide but left a suicide note stating that her suicide was due to depression and unrelated to the study and that her OCD was still improved. Still, because DBS is often a "last-resort" effort for patients to recapture their "lost" lives, the desperation of patients who qualify for DBS trials may create a risk for suicide. It may also complicate the science of testing efficacy by creating fertile ground for placebo responses. However, placebo effects are rarely long lasting, and OCD is fairly resistant to them.[323]

Greenberg *et al.*[324] reported their results from a collaborative study at Brown Medical School and the Cleveland Clinic on the long-term outcomes of 10 treatment-resistant OCD patients receiving DBS bilaterally in the ventral capsule/ventral striatum. There was a significant positive effect of long-term stimulation (36 months postoperatively) in terms of improved Y-BOCS and Global Assessment of Functioning scores, depression, anxiety, self-care, independent living, and work, school, and social functioning. Four of the 8 patients who reached the 36-month follow-up had a ≥35% decrease in their Y-BOCS scores, and Y-BOCS scores declined between 25% and 35% in 2 patients. Surgical adverse effects included 1 seizure, 1 superficial infection, and an asymptomatic hemorrhage. Psychiatric adverse effects included transient hypomanic symptoms and worsened depression and OCD when stimulator battery depletion interrupted DBS. Still, these results suggest that DBS can have positive long-term effects and is thus a promising treatment option for highly treatment-resistant OCD. A PET study of 6 OCD patients from this series found that acute high-frequency DBS of the ventral capsule/ventral striatum, compared to low-frequency DBS and stimulation-off conditions, increased activation (regional cerebral blood flow) of circuitry implicated in OCD, that is, in the OFC, anterior cingulate cortex, striatum, globus pallidus, and thalamus.[312]

Finally, 4 recent case reports of open-label DBS[325–328] support its efficacy in treatment-resistant OCD patients. In general, positive results were maintained in subsequent follow-up visits for up to 15 months with prolonged stimulation, and there were no serious adverse events. An intimate relationship between various neuropsychiatric circuits is suggested by some of these case studies examining DBS in individuals with comorbid disorders. For example, subthalamic nucleus stimulation in 2 PD patients with severe OCD significantly reduced their PD symptoms as well as their OCD symptoms; this was an unintentional benefit.[327] This implicates subcortical structures in the maintenance of OCD and a close connection between neuropsychiatric circuits. Another patient with comorbid major depression and OCD had remission of depression at 6 months and of OCD at 12 to 15 months after DBS of the ventral caudate nucleus, and the benefits continued for the 15-month follow-up period.[328] Recent exploration of DBS of the subgenual cingulate for treatment-resistant depression has also yielded positive results.[329]

Summary of DBS

In sum, DBS has potential value for treating refractory OCD, but additional development work is needed. The results of several DBS trials for OCD are difficult to generalize because of variations in the targets for stimulation and in methodology. The studies thus far have been too small to reach any significant conclusions about efficacy. A more definitive test of the efficacy and tolerability of DBS will require larger, blinded, controlled trials with select patients who meet rigorous inclusion criteria and comprehensive and continuous evaluation of outcome. Fortunately, this relatively nondestructive, reversible intervention makes well-designed controlled trials possible. Therefore, given the preliminary promising results in treatment-resistant OCD and the procedure's reversibility and adjustability in comparison with ablative neurosurgery, DBS deserves further investigation as it may be an effective therapeutic alternative in select patients with severe, treatment-resistant OCD. However, although DBS clearly has shown some promise and was generally well tolerated in most studies, it is still an invasive procedure with the associated risk of adverse events, both those related to the surgical procedure itself such as hemorrhage, seizure, and infection at the surgical incision site[330] and psychiatric adverse events such as DBS-induced acute mood elevation or OCD symptom worsening if DBS is interrupted. Therefore, in the clinic, DBS for psychiatric illness should be used conservatively, and a team of interdisciplinary experts in patient selection, implantation, stimulation, and long-term patient monitoring should be used. Counseling after DBS surgery is also recommended to help patients implement their improvements in daily life.[316]

The selection of the appropriate target is key for effective DBS for OCD.[112] The DBS targets for OCD reported in the literature were selected empirically and/or from an understanding of the presumed pathophysiology of OCD. Several targets have been proposed, including the NA, anterior limb of the internal capsule, ventral portions of the striatum, and ventral caudate nucleus. The differential success of these targets suggests that each may have a distinct role in OCD-related cognitive loops and/or may reflect differences in the study samples. However, the close anatomic relationship between some of these targets and the use of propagating electrical impulses makes overlap likely. The need for high-amplitude stimulation in these procedures suggests that the target is not well circumscribed or that current spread beyond the target is important to elicit the effect. Additional work is needed to identify more specific target regions that are tied to the pathophysiology of OCD. Promising results, similar to those of ablative surgical procedures, have been found with stimulation of the anterior limb of the internal capsule and of the NA. However, there is currently no consensus in the literature on a "gold-standard" DBS target or stimulation amplitude (effective stimulation has ranged from 2 to 10.5 V) for the treatment of OCD. This creates problems for the comparison of studies and the interpretation of results.

The focal and adjustable nature of DBS makes it ideal for offering insights into the neurocircuitry involved in OCD.[324] For example, some have suggested that high-frequency DBS in the internal capsule blocks hyperactive circuits of OCD and is related to the observed clinical benefits. Neuroimaging studies[312] are starting to investigate exactly how DBS affects these circuits. More accurate, empirically grounded anatomic localization may allow the use of less stimulation to achieve the same clinical benefits.[331] Surgically targeting new regions thought to be involved in OCD would offer patients more therapeutic options.[112] More research could also be aimed at the development of implantable, rechargeable batteries with longer lives and better designed electrodes and stimulation devices.[331] In sum, DBS is an important advance in patient care that may significantly improve the quality of life of a number of patients who have exhausted almost all other treatment options.

Treatment-resistant and intractable OCD is still an important challenge for researchers and clinicians. Adjunctive antipsychotics and other promising somatic treatments such as TMS and DBS need more investigation to establish efficacy. In recent years, brain stimulation methods have been defining a "third path" of treatment outside of traditional pharmacological and psychotherapy.[270] Treatment studies with brain stimulation techniques (TMS, ECT, and DBS) in severe, treatment-refractory or intractable OCD patients have shown varying levels of efficacy depending on the technique. In general, the level of confidence in these results is restricted by design limitations. Studies exploring these techniques for OCD treatment are limited by small sample sizes and scarcity of double-blind trials, and none of the alternative interventions reviewed here are FDA-approved for the treatment of OCD. However, given the promising findings thus far, reversibility, noninvasiveness or minimal invasiveness, and possibility of double-blind trials, additional research should be conducted with TMS and DBS, which may refine these techniques and establish their efficacy. Larger controlled trials with select patients meeting strict inclusion criteria are needed to more definitively test the efficacy and tolerability of DBS as a potential neurosurgical treatment alternative for refractory OCD. In addition to the potential of DBS to relieve suffering and improve the lives of people with OCD, given its focality and adjustability, it can also be used to investigate the neurocircuitry involved in OCD.

FUTURE DIRECTIONS OF TREATMENT

Current treatment approaches can alleviate OCD symptoms in some patients, but additional research is needed to enhance existing treatments and develop safer and more effective therapeutic options. Studies are needed to explore whether modifications to treatment regimens can increase the number of responders and the rapidity, degree, and endurance of response. For example, studies could investigate whether faster titration or higher doses of SRIs result in more rapid or greater symptom relief. Furthermore, a number of medications other than SRIs have shown some efficacy for OCD in preliminary studies, either as monotherapy or as augmentation (see Table 7). As previously discussed, certain anticonvulsants have been found to help some OCD patients either as monotherapy or as augmentation agents. In addition, initial studies on the use of mirtazapine,[332,333] phenelzine,[334] pindolol,[335,336] ondansetron,[337] inositol,[338,339] riluzole,[84] D-amphetamine,[340] and tramadol[341] have shown some promising results as well. However, further study of the efficacy of these medications in larger randomized, controlled trials or within the context of proposed animal models would further enhance the understanding of OCD treatment as well as its underlying neurobiology. Additional studies of psychosocial therapies and the various combinations of these therapies with medication are also needed. Furthermore, studies are needed with all treatment approaches to OCD that will help target specific treatments to individual patients. For example, the most effective treatment may vary according to the OCD symptom type (eg, contamination, hoarding, and mental compulsions) or comorbid condition (eg, tics, schizotypal personality disorder, and depression).

Table 7. *Novel Pharmacotherapies That May Be Effective for Refractory Obsessive-Compulsive Disorder.*

Drug	Class
Phenelzine	Monoamine oxidase inhibitor
Pindolol	Nonselective beta-blocker
Mirtazapine	Piperazino-azepine
Morphine sulfate	Opioid analgesic
Tramadol	Central acting analgesic
Valproate	Carboxylic acid derivative (anticonvulsant)
Oxcarbazapine	Dibenzazepine (anticonvulsant)
Carbamazepine	Carboxamide (anticonvulsant)
Gabapentin	GABA analog (anticonvulsant)
Topiramate	Sulfamate-substituted monosaccharide antiepileptic (anticonvulsant)
Ondansetron	5-HT3 antagonist
D-Amphetamine	Sympathomimetic amine
Riluzole	Benzothiazole
Lithium	Antimanic agent
Inositol	Carbocyclic polyol

Abbreviations: GABA, γ -aminobutyric acid; 5-HT3, 5-hydroxytryptamine receptor 3.

Improved understanding of the neurobiological basis of OCD and of treatment response will help in the development of new treatment strategies. For example, identification of susceptibility genes for OCD could provide insight into the pathogenesis of OCD, establish a way of identifying those at risk for developing the disorder, and lead to the development of more targeted and effective treatments. Advances in pharmacogenetics, such as gene chip techniques, could help establish new neurochemical targets for treatment and help to predict responses to or side effects of specific medications or classes of medications. A single-microarray assay, DNA microarray, or gene chip could allow for comprehensive genotyping of thousands of variants in genes involved in drug metabolism, excretion, and transport. Furthermore, because of genetic similarity, the response of first-degree relatives to certain medications may predict the patient's response. Perhaps this can be explored in future research as well because no definitive data are currently available. Other markers (eg, neurocognitive tests, blood, electroencephalogram, and functional and structural neuroimaging) may also aid in understanding the underlying neurobiology of OCD and in identifying early effects that may predict individual treatment response. For example, some small PET studies[206,207] have shown correlations between glucose metabolic rates in specific brain areas and OCD symptom response to pharmacological or behavioral therapy.

To date, most attention has been focused on nonhuman primate models of OCD.[342] Neuroimaging techniques could be used to look at component features of OCD, such as anxiety, compulsions, and maladaptive thoughts, to help formulate a model of OCD specific to the symptoms of the disorder in humans.[69] Many neuroimaging studies have investigated only a limited number of brain areas, and this has biased studies toward verification of well-established models.[66] Future neuroimaging studies should more thoroughly examine the brain, using, for example, whole brain–based imaging techniques to determine whether abnormalities associated with OCD are limited to the corticostriatothalamocortical circuitry. For example, several recent studies have implicated the amygdala[46] and parietal cortices[49,73] in OCD, prompting a reconceptualization of current theories of the pathophysiology of OCD. Future studies could also examine the circuitry of highly comorbid disorders with OCD (eg, attention-deficit hyperactivity disorder, other anxiety disorders, and depression) for similarities and differences to provide further insight into the etiology of OCD. Imaging studies should also compare OCD to other commonly comorbid disorders in order to establish the specificity of imaging findings in OCD.[342]

Genetic models are useful in identifying specific genes and neuroanatomical pathways important in OCD (see Finn *et al.*[343]). Although candidate gene studies are promising, they are not yet associated with linkage regions, except for the glutamate transporter gene *SLC1A1* in 9p24.[344] A number of OCD genetic association studies (case-control and family-based) have implicated genes encoding components within the 5-hydroxytryptamine receptor (5-HT) and dopaminergic systems.[127] However, additional innovative research is needed to unravel the precise genetic components of OCD. For example, breaking down OCD symptoms into various dimensions that are continuous with the normal population could help to disentangle the complex inheritance of OCD.[345] Furthermore, because there are advantages of gene-based approaches versus single-nucleotide polymorphism–based approaches,[346] future studies could employ more complete assessments of candidate genes, possibly using haplotype blocks that span larger regions.[127]

Animal models have proven invaluable in OCD research, confirming, for example, the importance of the 5-HT[157,347,348] and dopamine systems[169,349] in the neurobiology and treatment of OCD. Mutant rodent models have provided important information about how genes and the environment interact to affect OCD behaviors via multiple neuropharmacological pathways.[343] To date, a number of specific genetically engineered animal models of OCD have been developed. However, their predictive validity is limited by the lack of reports of the effects of pharmacological treatments in these models.[153] Despite this, there are several promising genes. For example, mice with genetic deletion of SAPAP3 exhibit increased anxiety and compulsive grooming behavior.[126] SAPAP3, a postsynaptic scaffolding protein at excitatory synapses, appears to play a critical role at corticostriatal synapses. This, in conjunction with genetic association studies of OCD in humans implicating genes involved in glutamatergic neurotransmission,[115,124,125] suggests that defects in excitatory synaptic transmission in the corticostriatal circuit may contribute to the pathogenesis of OCD. *Hoxb8* mutant mice also show increased grooming behavior,[170] and the *Hoxb8* gene is expressed in the OFC, anterior cingulate, striatum, and limbic system, all of which are implicated in OCD.

Although transgenic models have great potential, further investigation is needed to determine whether the subsequent OCD-like behaviors are a direct result of the manipulated gene or are caused by other targets downstream of the gene. We need to know if the gene under investigation is in fact positively associated with the underlying pathology of OCD. Expression array technologies could be employed, which may implicate novel candidates and neurobiological pathways.

The results from genetic studies should be incorporated with clinical and epidemiologic factors to elucidate the cause of OCD. For example, because SRIs are not always effective for OCD,[350] a lack of effect in an animal model may reflect an unknown neurobiological basis for compulsive behavior in a subgroup of SRI-refractory patients. Furthermore, separating the afflicted (ie, working with animals that show greater behavioral change in a model and/or after drug treatment) could be beneficial. For a more comprehensive understanding of OCD, multiple animal models will be needed to explore the various aspects of OCD and provide convergent validation of research findings. The development of valid animal models for OCD will increase our understanding of the neural basis of OCD and aid in the development and identification of new therapeutic strategies. Ultimately, animal models can provide new molecular targets for the treatment of OCD in human populations, a key goal of transitional research.

Genetically engineered animal models will be increasingly valuable, in conjunction with new technologies such as RNA interference, gene-chip microarrays, and advanced proteomics, in helping to further our understanding of OCD. Noninvasive neuroimaging (eg, small-animal single-photon emission computed tomography) for exploring the neuroanatomic basis of OCD also offers an exciting future direction, especially if combined with pharmacologic models. Neuroimaging and cross-species translational research of the physiological and anatomical pathways implicated in the pathophysiology and treatment response in OCD will advance our understanding of the neural basis of OCD and lead to more targeted and effective treatment options.

DISCLOSURES

Eric Hollander, MD had research grants from Solvay, Janssen, Bristol-Meyers Squibb, and GlaxoSmithKline in the past.

REFERENCES

1. Hay P, Sachdev P, Cumming S, et al. Treatment of obsessive-compulsive disorder by psychosurgery. *Acta Psychiatr Scand* 1993; 87: 197–207.
2. DuPont RL, Rice DP, Shiraki S, et al. Economic costs of obsessive-compulsive disorder. *Med Interface* 1995; 8: 102–109.
3. Steketee G. Disability and family burden in obsessive-compulsive disorder. *Can J Psychiatry* 1997; 42: 919–928.
4. Stein D, Allen A, Bobes J, et al. Quality of life in obsessive compulsive disorder. *CNS Spectr* 2000; 5(suppl 4): 37–39.
5. Stein DJ. Obsessive-compulsive disorder. *Lancet* 2002; 360: 397–405.
6. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Text Revision. Washington, DC: American Psychiatric Association; 2000.
7. Goodman WK. Obsessive-compulsive disorder: diagnosis and treatment. *J Clin Psychiatry* 1999; 60(suppl 18): 27–32.
8. Karno M, Golding JM, Sorenson SB, et al. The epidemiology of obsessive-compulsive disorder in five US communities. *Arch Gen Psychiatry* 1988; 45: 1094–1099.
9. Weissman MM, Bland RC, Canino GJ, et al., for the Cross National Collaborative Group. The cross national epidemiology of obsessive compulsive disorder. *J Clin Psychiatry* 1994; 55(suppl): 5–10.
10. Sasson Y, Zohar J, Chopra M, et al. Epidemiology of obsessive-compulsive disorder: a world view. *J Clin Psychiatry* 1997; 58(suppl 12): 7–10.
11. Torres AR, Prince MJ, Bebbington PE, et al. Obsessive-compulsive disorder: prevalence, comorbidity, impact, and help-seeking in the British National Psychiatric Morbidity Survey of 2000. *Am J Psychiatry* 2006; 163: 1978–1985.
12. Robins LN, Helzer JE, Weissman MM, et al. Lifetime prevalence of specific psychiatric disorders in three sites. *Arch Gen Psychiatry* 1984; 41: 949–958.
13. Stein MB, Forde DR, Anderson G, et al. Obsessive-compulsive disorder in the community: an epidemiologic survey with clinical reappraisal. *Am J Psychiatry* 1997; 154: 1120–1126.
14. Skoog G, Skoog I. A 40-year follow-up of patients with obsessive-compulsive disorder [see comments]. *Arch Gen Psychiatry* 1999; 56: 121–127.
15. Welkowitz LA, Struening EL, Pittman J, et al. Obsessive-compulsive disorder and comorbid anxiety problems in a national anxiety screening sample. *J Anxiety Disord* 2000; 14: 471–482.
16. Nestadt G, Addington A, Samuels J, et al. The identification of OCD-related subgroups based on comorbidity. *Biol Psychiatry* 2003; 53: 914–920.
17. Biederman J, Petty C, Faraone SV, et al. Moderating effects of major depression on patterns of comorbidity in patients with panic disorder. *Psychiatry Res* 2004; 126: 143–149.
18. LaSalle VH, Cromer KR, Nelson KN, et al. Diagnostic interview assessed neuropsychiatric disorder comorbidity in 334 individuals with obsessive-compulsive disorder. *Depress Anxiety* 2004; 19: 163–173.
19. Apter A, Horesh N, Gothelf D, et al. Depression and suicidal behavior in adolescent inpatients with obsessive compulsive disorder. *J Affect Disord* 2003; 75: 181–189.
20. Abramowitz JS, Schwartz SA, Moore KM, et al. Obsessive-compulsive symptoms in pregnancy and the puerperium: a review of the literature. *J Anxiety Disord* 2003; 17: 461–478.
21. Abramowitz JS. Treatment of obsessive-compulsive disorder in patients who have comorbid major depression. *J Clin Psychol* 2004; 60: 1133–1141.
22. Fals-Stewart W, Angarano K. Obsessive-compulsive disorder among patients entering substance abuse treatment. Prevalence and accuracy of diagnosis. *J Nerv Ment Dis* 1994; 182: 715–719.

DOI:10.1002/MSJ

23. Abram KM, Teplin LA, McClelland GM, et al. Comorbid psychiatric disorders in youth in juvenile detention. *Arch Gen Psychiatry* 2003; 60: 1097–1108.
24. Bakken K, Landheim AS, Vaglum P. Primary and secondary substance misusers: do they differ in substance-induced and substance-independent mental disorders? *Alcohol Alcohol* 2003; 38: 54–59.
25. Jordan J, Joyce PR, Carter FA, et al. Anxiety and psychoactive substance use disorder comorbidity in anorexia nervosa or depression. *Int J Eat Disord* 2003; 34: 211–219.
26. Frare F, Perugi G, Ruffolo G, et al. Obsessive-compulsive disorder and body dysmorphic disorder: a comparison of clinical features. *Eur Psychiatry* 2004; 19: 292–298.
27. Koran LM, Hanna GL, Hollander E, et al. Practice guideline for the treatment of patients with obsessive-compulsive disorder. *Am J Psychiatry* 2007; 164: 5–53.
28. Swedo S, Snider L. In: Nestler EJ, Charney DS, eds. Neurobiology of Mental Illness. 2nd Editions. The Neurobiology and Treatment of Obsessive-Compulsive Disorder. New York, NY: Oxford University Press; 2004; 628–638.
29. Graybiel AM, Rauch SL. Toward a neurobiology of obsessive-compulsive disorder. *Neuron* 2000; 28: 343–347.
30. Aouizerate B, Guehl D, Cuny E, et al. Pathophysiology of obsessive-compulsive disorder: a necessary link between phenomenology, neuropsychology, imagery and physiology. *Prog Neurobiol* 2004; 72: 195–221.
31. Saxena S, Bota RG, Brody AL. Brain-behavior relationships in obsessive-compulsive disorder. *Semin Clin Neuropsychiatry* 2001; 6: 82–101.
32. Alexander GE, DeLong MR, Strick PL. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu Rev Neurosci* 1986; 9: 357–381.
33. Lawrence A, Sahakian B, Robbins TW. Cognitive functions and corticostriatal circuits: insights from Huntington's disease. *Trends Cogn Sci* 1998; 2: 379–388.
34. Rapoport JL, Wise SP. Obsessive-compulsive disorder: evidence for basal ganglia dysfunction. *Psychopharmacol Bull* 1988; 24: 380–384.
35. Saxena S. Neuroimaging and the pathophysiology of obsessive compulsive disorder (OCD). In: Fu CHY, Senior C, Russell T, Weinberger DR, eds. Neuroimaging in Psychiatry. London, United Kingdom: Martin Dunitz; 2003.
36. Menzies L, Achard S, Chamberlain SR, et al. Neurocognitive endophenotypes of obsessive-compulsive disorder. *Brain* 2007; 130: 3223–3236.
37. Rolls ET. The functions of the orbitofrontal cortex. *Brain Cogn* 2004; 55: 11–29.
38. Berlin HA, Rolls ET, Kischka U. Impulsivity, time perception, emotion and reinforcement sensitivity in patients with orbitofrontal cortex lesions. *Brain* 2004; 127: 1108–1126.
39. Chamberlain SR, Blackwell AD, Fineberg NA, et al. The neuropsychology of obsessive compulsive disorder: the importance of failures in cognitive and behavioural inhibition as candidate endophenotypic markers. *Neurosci Biobehav Rev* 2005; 29: 399–419.
40. Tallis F. The neuropsychology of obsessive-compulsive disorder: a review and consideration of clinical implications. *Br J Clin Psychol* 1997; 36(pt 1): 3–20.
41. Robinson D, Wu H, Munne RA, et al. Reduced caudate nucleus volume in obsessive-compulsive disorder. *Arch Gen Psychiatry* 1995; 52: 393–398.
42. Scarone S, Colombo C, Livian S, et al. Increased right caudate nucleus size in obsessive-compulsive disorder: detection with magnetic resonance imaging. *Psychiatry Res* 1992; 45: 115–121.
43. Aylward EH, Harris GJ, Hoehn-Saric R, et al. Normal caudate nucleus in obsessive-compulsive disorder assessed by quantitative neuroimaging. *Arch Gen Psychiatry* 1996; 53: 577–584.
44. Jenike MA, Breiter HC, Baer L, et al. Cerebral structural abnormalities in obsessive-compulsive disorder. A quantitative morphometric magnetic resonance imaging study. *Arch Gen Psychiatry* 1996; 53: 625–632.
45. Rosenberg DR, Keshavan MS. A.E. Bennett Research Award. Toward a neurodevelopmental model of obsessive-compulsive disorder. *Biol Psychiatry* 1998; 43: 623–640.
46. Szeszko PR, Robinson D, Alvir JM, et al. Orbital frontal and amygdala volume reductions in obsessive-compulsive disorder. *Arch Gen Psychiatry* 1999; 56: 913–919.
47. Szeszko PR, MacMillan S, McMeniman M, et al. Brain structural abnormalities in psychotropic drug-naive pediatric patients with obsessive-compulsive disorder. *Am J Psychiatry* 2004; 161: 1049–1056.
48. Gilbert AR, Moore GJ, Keshavan MS, et al. Decrease in thalamic volumes of pediatric patients with obsessive-compulsive disorder who are taking paroxetine. *Arch Gen Psychiatry* 2000; 57: 449–456.
49. Kim JJ, Lee MC, Kim J, et al. Grey matter abnormalities in obsessive-compulsive disorder: statistical parametric mapping of segmented magnetic resonance images. *Br J Psychiatry* 2001; 179: 330–334.
50. Kwon JS, Shin YW, Kim CW, et al. Similarity and disparity of obsessive-compulsive disorder and schizophrenia in MR volumetric abnormalities of the hippocampus-amygdala complex. *J Neurol Neurosurg Psychiatry* 2003; 74: 962–964.
51. Choi JS, Kang DH, Kim JJ, et al. Left anterior subregion of orbitofrontal cortex volume reduction and impaired organizational strategies in obsessive-compulsive disorder. *J Psychiatr Res* 2004; 38: 193–199.
52. Choi JS, Kim HS, Yoo SY, et al. Morphometric alterations of anterior superior temporal cortex in obsessive-compulsive disorder. *Depress Anxiety* 2006; 23: 290–296.
53. Kang DH, Kim JJ, Choi JS, et al. Volumetric investigation of the frontal-subcortical circuitry in patients with obsessive-compulsive disorder. *J Neuropsychiatry Clin Neurosci* 2004; 16: 342–349.
54. Atmaca M, Yildirim H, Ozdemir H, et al. Volumetric MRI study of key brain regions implicated in obsessive-compulsive disorder. *Prog Neuropsychopharmacol Biol Psychiatry* 2007; 31: 46–52.
55. Rauch SL, Jenike MA, Alpert NM, et al. Regional cerebral blood flow measured during symptom provocation in obsessive-compulsive disorder using oxygen 15-labeled carbon dioxide and positron emission tomography. *Arch Gen Psychiatry* 1994; 51: 62–70.

56. McKay D, Abramowitz JS, Calamari JE, et al. A critical evaluation of obsessive-compulsive disorder subtypes: symptoms versus mechanisms. *Clin Psychol Rev* 2004; 24: 283–313.
57. Nordahl TE, Benkelfat C, Semple WE, et al. Cerebral glucose metabolic rates in obsessive compulsive disorder. *Neuropsychopharmacology* 1989; 2: 23–28.
58. Baxter LR. Brain imaging as a tool in establishing a theory of brain pathology in obsessive compulsive disorder. *J Clin Psychiatry* 1990; 51(():suppl) 22–25.
59. McGuire PK, Bench CJ, Frith CD, et al. Functional anatomy of obsessive-compulsive phenomena. *Br J Psychiatry* 1994; 164: 459–468.
60. Phillips ML, Marks IM, Senior C, et al. A differential neural response in obsessive-compulsive disorder patients with washing compared with checking symptoms to disgust. *Psychol Med* 2000; 30: 1037–1050.
61. Lacerda AL, Dalgalarrondo P, Caetano D, et al. Elevated thalamic and prefrontal regional cerebral blood flow in obsessive-compulsive disorder: a SPECT study. *Psychiatry Res* 2003; 123: 125–134.
62. Mataix-Cols D, Wooderson S, Lawrence N, et al. Distinct neural correlates of washing, checking, and hoarding symptom dimensions in obsessive-compulsive disorder. *Arch Gen Psychiatry* 2004; 61: 564–576.
63. Saxena S, Brody AL, Maidment KM, et al. Cerebral glucose metabolism in obsessive-compulsive hoarding. *Am J Psychiatry* 2004; 161: 1038–1048.
64. Whiteside SP, Port JD, Abramowitz JS. A meta-analysis of functional neuroimaging in obsessive-compulsive disorder. *Psychiatry Res* 2004; 132: 69–79.
65. Saxena S, Rauch SL. Functional neuroimaging and the neuroanatomy of obsessive-compulsive disorder. *Psychiatr Clin N Am* 2000; 23: 563–586.
66. Rauch SL. Neuroimaging and neurocircuitry models pertaining to the neurosurgical treatment of psychiatric disorders. *Neurosurg Clin N Am* 2003; 14: 213–223, vii–viii.
67. Saxena S, Brody AL, Schwartz JM, et al. Neuroimaging and frontal-subcortical circuitry in obsessive-compulsive disorder. *Br J Psychiatry Suppl* 1998; 35: 26–37.
68. Saxena S, Brody AL, Maidment KM, et al. Localized orbitofrontal and subcortical metabolic changes and predictors of response to paroxetine treatment in obsessive-compulsive disorder. *Neuropsychopharmacology* 1999; 21: 683–693.
69. Friedlander L, Desrocher M. Neuroimaging studies of obsessive-compulsive disorder in adults and children. *Clin Psychol Rev* 2006; 26: 32–49.
70. Bullmore ET, Suckling J, Overmeyer S, et al. Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Trans Med Imaging* 1999; 18: 32–42.
71. Ashburner J, Friston KJ. Voxel-based morphometry–the methods. *Neuroimage* 2000; 11: 805–821.
72. Pujol J, Soriano-Mas C, Alonso P, et al. Mapping structural brain alterations in obsessive-compulsive disorder. *Arch Gen Psychiatry* 2004; 61: 720–730.
73. Valente AA Jr, Miguel EC, Castro CC, et al. Regional gray matter abnormalities in obsessive-compulsive disorder: a voxel-based morphometry study. *Biol Psychiatry* 2005; 58: 479–487.
74. Husted DS, Shapira NA, Goodman WK. Obsessive-compulsive disorder. In: Tarazi FI, Schetz JA, eds. *Neurological and Psychiatric Disorders: From Bench to Bedside*. Totowa, NJ: Humana Press; 2005; 171–188.
75. Zohar J, Mueller EA, Insel TR, et al. Serotonergic responsivity in obsessive-compulsive disorder. Comparison of patients and healthy controls. *Arch Gen Psychiatry* 1987; 44: 946–951.
76. Hollander E, DeCaria CM, Nitescu A, et al. Serotonergic function in obsessive-compulsive disorder. Behavioral and neuroendocrine responses to oral m-chlorophenylpiperazine and fenfluramine in patients and healthy volunteers. *Arch Gen Psychiatry* 1992; 49: 21–28.
77. Stein DJ, Van Heerden B, Wessels CJ, et al. Single photon emission computed tomography of the brain with Tc-99m HMPAO during sumatriptan challenge in obsessive-compulsive disorder: investigating the functional role of the serotonin auto-receptor. *Prog Neuropsychopharmacol Biol Psychiatry* 1999; 23: 1079–1099.
78. Adams KH, Hansen ES, Pinborg LH, et al. Patients with obsessive-compulsive disorder have increased 5-HT2A receptor binding in the caudate nuclei. *Int J Neuropsychopharmacol* 2005; 8: 391–401.
79. Denys D, Zohar J, Westenberg HG. The role of dopamine in obsessive-compulsive disorder: preclinical and clinical evidence. *J Clin Psychiatry* 2004; 65(suppl 14): 11–17.
80. van der Wee NJ, Stevens H, Hardeman JA, et al. Enhanced dopamine transporter density in psychotropic-naive patients with obsessive-compulsive disorder shown by [123I]beta-CIT SPECT. *Am J Psychiatry* 2004; 161: 2201–2206.
81. Szechtman H, Sulis W, Eilam D. Quinpirole induces compulsive checking behavior in rats: a potential animal model of obsessive-compulsive disorder (OCD). *Behav Neurosci* 1998; 112: 1475–1485.
82. Pittenger C, Krystal JH, Coric V. Glutamate-modulating drugs as novel pharmacotherapeutic agents in the treatment of obsessive-compulsive disorder. *NeuroRx* 2006; 3: 69–81.
83. McGrath MJ, Campbell KM, Parks CR, et al. Glutamatergic drugs exacerbate symptomatic behavior in a transgenic model of comorbid Tourette's syndrome and obsessive-compulsive disorder. *Brain Res* 2000; 877: 23–30.
84. Coric V, Taskiran S, Pittenger C, et al. Riluzole augmentation in treatment-resistant obsessive-compulsive disorder: an open-label trial. *Biol Psychiatry* 2005; 58: 424–428.
85. Bhattacharyya S, Chakraborty K. Glutamatergic dysfunction–newer targets for anti-obsessional drugs. *Recent Patents CNS Drug Discov* 2007; 2: 47–55.
86. Van Ameringen M, Mancini C, Patterson B, et al. Topiramate augmentation in treatment-resistant obsessive-compulsive disorder: a retrospective, open-label case series. *Depress Anxiety* 2006; 23: 1–5.
87. Rubio G, Jimenez-Arriero MA, Martinez-Gras I, et al. The effects of topiramate adjunctive treatment added to antidepressants in patients with resistant obsessive-compulsive disorder. *J Clin Psychopharmacol* 2006; 26: 341–344.

88. Hollander E, Dell'Osso B. Topiramate plus paroxetine in treatment-resistant obsessive-compulsive disorder. *Int Clin Psychopharmacol* 2006; 21: 189–191.
89. Coric V, Milanovic S, Wasylink S, et al. Beneficial effects of the antiglutamatergic agent riluzole in a patient diagnosed with obsessive-compulsive disorder and major depressive disorder. *Psychopharmacology (Berl)* 2003; 167: 219–220.
90. Lafleur DL, Pittenger C, Kelmendi B, et al. N-acetylcysteine augmentation in serotonin reuptake inhibitor refractory obsessive-compulsive disorder. *Psychopharmacology (Berl)* 2006; 184: 254–256.
91. Berlin HA, Koran L, Jenike M, et al. Double-blind, placebo-controlled trial of topiramate augmentation in treatment-resistant obsessive-compulsive disorder in preperation for neuropsychopharmacology.
92. Nestadt G, Samuels J, Riddle M, et al. A family study of obsessive-compulsive disorder. *Arch Gen Psychiatry* 2000; 57: 358–363.
93. Inouye E. Similar and dissimilar manifestations of obsessive-compulsive neuroses in monozygotic twins. *Am J Psychiatry* 1965; 121: 1171–1175.
94. Hettema JM, Neale MC, Kendler KS. A review and meta-analysis of the genetic epidemiology of anxiety disorders. *Am J Psychiatry* 2001; 158: 1568–1578.
95. Carey G II. Twin and family studies of anxiety, phobic and obsessive disorders. In: Klein DF, Rabkin JG, eds. *Anxiety: New Research and Challenging Concepts*. New York, NY: Raven Press; 1981: 117–136.
96. Clifford CA, Murray RM, Fulker DW. Genetic and environmental influences on obsessional traits and symptoms. *Psychol Med* 1984; 14: 791–800.
97. Jonnal AH, Gardner CO, Prescott CA, et al. Obsessive and compulsive symptoms in a general population sample of female twins. *Am J Med Genet* 2000; 96: 791–796.
98. Pauls DL, Alsobrook JP II, Goodman W, et al. A family study of obsessive-compulsive disorder. *Am J Psychiatry* 1995; 152: 76–84.
99. Bellodi L, Sciuto G, Diaferia G, et al. Psychiatric disorders in the families of patients with obsessive-compulsive disorder. *Psychiatry Res* 1992; 42: 111–120.
100. do Rosario-Campos MC, Leckman JF, Curi M, et al. A family study of early-onset obsessive-compulsive disorder. *Am J Med Genet B Neuropsychiatr Genet* 2005; 136: 92–97.
101. Hanna GL, Himle JA, Curtis GC, et al. A family study of obsessive-compulsive disorder with pediatric probands. *Am J Med Genet B Neuropsychiatr Genet* 2005; 134: 13–19.
102. Alsobrook IJ, Leckman JF, Goodman WK, et al. Segregation analysis of obsessive-compulsive disorder using symptom-based factor scores. *Am J Med Genet* 1999; 88: 669–675.
103. Pauls DL, Alsobrook JP II. The inheritance of obsessive-compulsive disorder. *Child Adolesc Psychiatr Clin N Am* 1999; 8: 481–496.
104. Hanna GL, Veenstra-VanderWeele J, Cox NJ, et al. Genome-wide linkage analysis of families with obsessive-compulsive disorder ascertained through pediatric probands. *Am J Med Genet* 2002; 114: 541–552.
105. Shugart YY, Samuels J, Willour VL, et al. Genome-wide linkage scan for obsessive-compulsive disorder: evidence for susceptibility loci on chromosomes 3q, 7p, 1q, 15q, and 6q. *Mol Psychiatry* 2006; 11: 763–770.
106. Pitman RK, Green RC, Jenike MA, et al. Clinical comparison of Tourette's disorder and obsessive-compulsive disorder. *Am J Psychiatry* 1987; 144: 1166–1171.
107. Leonard HL, Lenane MC, Swedo SE, et al. Tics and Tourette's disorder: a 2- to 7-year follow-up of 54 obsessive-compulsive children. *Am J Psychiatry* 1992; 149: 1244–1251.
108. Rapoport JL, Inoff-Germain G. Treatment of obsessive-compulsive disorder in children and adolescents. *J Child Psychol Psychiatry* 2000; 41: 419–431.
109. Grados MA, Riddle MA, Samuels JF, et al. The familial phenotype of obsessive-compulsive disorder in relation to tic disorders: the Hopkins OCD family study. *Biol Psychiatry* 2001; 50: 559–565.
110. Pauls DL, Leckman JF. The inheritance of Gilles de la Tourette's syndrome and associated behaviors. Evidence for autosomal dominant transmission. *N Engl J Med* 1986; 315: 993–997.
111. Walkup JT, LaBuda MC, Singer HS, et al. Family study and segregation analysis of Tourette syndrome: evidence for a mixed model of inheritance. *Am J Hum Genet* 1996; 59: 684–693.
112. Lipsman N, Neimat JS, Lozano AM. Deep brain stimulation for treatment-refractory obsessive-compulsive disorder: the search for a valid target. *Neurosurgery* 2007; 61: 1–11; discussion 11–13.
113. Willour VL, Yao Shugart Y, Samuels J, et al. Replication study supports evidence for linkage to 9p24 in obsessive-compulsive disorder. *Am J Hum Genet* 2004; 75: 508–513.
114. Veenstra-VanderWeele J, Kim SJ, Gonen D, et al. Genomic organization of the SLC1A1/EAAC1 gene and mutation screening in early-onset obsessive-compulsive disorder. *Mol Psychiatry* 2001; 6: 160–167.
115. Arnold PD, Rosenberg DR, Mundo E, et al. Association of a glutamate (NMDA) subunit receptor gene (GRIN2B) with obsessive-compulsive disorder: a preliminary study. *Psychopharmacology (Berl)* 2004; 174: 530–538.
116. Chabane N, Millet B, Delorme R, et al. Lack of evidence for association between serotonin transporter gene (5-HTTLPR) and obsessive-compulsive disorder by case control and family association study in humans. *Neurosci Lett* 2004; 363: 154–156.
117. Hall D, Dhilla A, Charalambous A, et al. Sequence variants of the brain-derived neurotrophic factor (BDNF) gene are strongly associated with obsessive-compulsive disorder. *Am J Hum Genet* 2003; 73: 370–376.
118. Millet B, Chabane N, Delorme R, et al. Association between the dopamine receptor D4 (DRD4) gene and obsessive-compulsive disorder. *Am J Med Genet B Neuropsychiatr Genet* 2003; 116: 55–59.
119. Mundo E, Richter MA, Zai G, et al. 5HT1Dbeta receptor gene implicated in the pathogenesis of obsessive-compulsive disorder: further evidence from a family-based association study. *Mol Psychiatry* 2002; 7: 805–809.
120. Zai G, Bezchlibnyk YB, Richter MA, et al. Myelin oligodendrocyte glycoprotein (MOG) gene is

associated with obsessive-compulsive disorder. *Am J Med Genet B Neuropsychiatr Genet* 2004; 129: 64–68.

121. Ozaki N, Goldman D, Kaye WH, et al. Serotonin transporter missense mutation associated with a complex neuropsychiatric phenotype. *Mol Psychiatry* 2003; 8: 933–936.
122. Billett EA, Richter MA, Sam F, et al. Investigation of dopamine system genes in obsessive-compulsive disorder. *Psychiatr Genet* 1998; 8: 163–169.
123. Baxter LR, Clark EC, Iqbal M, et al. Cortical-subcortical systems in the mediation of obsessive-compulsive disorder. In: Lichter DG, Cummings JL, eds. Frontal-Subcortical Circuits in Psychiatric and Neurological Disorders. New York, NY: Guilford Press; 2000; 207–230.
124. Arnold PD, Sicard T, Burroughs E, et al. Glutamate transporter gene SLC1A1 associated with obsessive-compulsive disorder. *Arch Gen Psychiatry* 2006; 63: 769–776.
125. Dickel DE, Veenstra-VanderWeele J, Cox NJ, et al. Association testing of the positional and functional candidate gene SLC1A1/EAAC1 in early-onset obsessive-compulsive disorder. *Arch Gen Psychiatry* 2006; 63: 778–785.
126. Welch JM, Lu J, Rodriguiz RM, et al. Cortico-striatal synaptic defects and OCD-like behaviours in Sapap3-mutant mice. *Nature* 2007; 448: 894–900.
127. Hemmings SM, Stein DJ. The current status of association studies in obsessive-compulsive disorder. *Psychiatr Clin N Am* 2006; 29: 411–444.
128. Gothelf D, Aharonovsky O, Horesh N, et al. Life events and personality factors in children and adolescents with obsessive-compulsive disorder and other anxiety disorders. *Compr Psychiatry* 2004; 45: 192–198.
129. Khanna S, Rajendra PN, Channabasavanna SM. Life events and onset of obsessive compulsive disorder. *Int J Soc Psychiatry* 1988; 34: 305–309.
130. de Silva P, Marks M. The role of traumatic experiences in the genesis of obsessive-compulsive disorder. *Behav Res Ther* 1999; 37: 941–951.
131. Lochner C, du Toit PL, Zungu-Dirwayi N, et al. Childhood trauma in obsessive-compulsive disorder, trichotillomania, and controls. *Depress Anxiety* 2002; 15: 66–68.
132. Neziroglu F, Anemone R, Yaryura-Tobias JA. Onset of obsessive-compulsive disorder in pregnancy. *Am J Psychiatry* 1992; 149: 947–950.
133. Sichel DA, Cohen LS, Dimmock JA, et al. Postpartum obsessive compulsive disorder: a case series. *J Clin Psychiatry* 1993; 54: 156–159.
134. Maina G, Albert U, Bogetto F, et al. Recent life events and obsessive-compulsive disorder (OCD): the role of pregnancy/delivery. *Psychiatry Res* 1999; 89: 49–58.
135. Lochner C, Hemmings SM, Kinnear CJ, et al. Gender in obsessive-compulsive disorder: clinical and genetic findings. *Eur Neuropsychopharmacol* 2004; 14: 105–113.
136. Williams KE, Koran LM. Obsessive-compulsive disorder in pregnancy, the puerperium, and the premenstruum. *J Clin Psychiatry* 1997; 58: 330–334; quiz 335–336.
137. Giulino L, Gammon P, Sullivan K, et al. Is parental report of upper respiratory infection at the onset of obsessive-compulsive disorder suggestive of pediatric autoimmune neuropsychiatric disorder associated with streptococcal infection? *J Child Adolesc Psychopharmacol* 2002; 12: 157–164.
138. Murphy TK, Sajid M, Soto O, et al. Detecting pediatric autoimmune neuropsychiatric disorders associated with streptococcus in children with obsessive-compulsive disorder and tics. *Biol Psychiatry* 2004; 55: 61–68.
139. Snider LA, Swedo SE. Childhood-onset obsessive-compulsive disorder and tic disorders: case report and literature review. *J Child Adolesc Psychopharmacol* 2003; 13(suppl 1): S81–S88.
140. Mell LK, Davis RL, Owens D. Association between streptococcal infection and obsessive-compulsive disorder, Tourette's syndrome, and tic disorder. *Pediatrics* 2005; 116: 56–60.
141. Dale RC, Heyman I. Post-streptococcal autoimmune psychiatric and movement disorders in children. *Br J Psychiatry* 2002; 181: 188–190.
142. Dale RC, Heyman I, Giovannoni G, et al. Incidence of anti-brain antibodies in children with obsessive-compulsive disorder. *Br J Psychiatry* 2005; 187: 314–319.
143. Swedo SE, Leonard HL, Garvey M, et al. Pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections: clinical description of the first 50 cases. *Am J Psychiatry* 1998; 155: 264–271.
144. Swedo SE, Leonard HL, Rapoport JL. The pediatric autoimmune neuropsychiatric disorders associated with streptococcal infection (PANDAS) subgroup: separating fact from fiction. *Pediatrics* 2004; 113: 907–911.
145. Luo F, Leckman JF, Katsovich L, et al. Prospective longitudinal study of children with tic disorders and/or obsessive-compulsive disorder: relationship of symptom exacerbations to newly acquired streptococcal infections. *Pediatrics* 2004; 113(:): e578–585.
146. Clark DA. Cognitive-Behavioral Therapy for OCD. New York, NY: Guilford Press; 2004.
147. Rachman S. A cognitive theory of obsessions. *Behav Res Ther* 1997; 35: 793–802.
148. Salkovskis PM, Richards C, Forrester E. The relationship between obsessional problems and intrusive thoughts. *Behav Cogn Psychother* 1995; 23: 281–299.
149. Wells A. Emotional Disorders and Metacognition: Innovative Cognitive Therapy. New York, NY: John Wiley & Sons; 2000.
150. Berle D, Starcevic V. Thought-action fusion: review of the literature and future directions. *Clin Psychol Rev* 2005; 25: 263–284.
151. Frost R, Steketee G. *Cognitive Approaches to Obsessions and Compulsions: Theory, Assessment, and Treatment*; Amsterdam, The Netherlands: Pergamon/Elsevier Science; 2002.
152. Joel D. Current animal models of obsessive compulsive disorder: a critical review. *Prog Neuropsychopharmacol Biol Psychiatry* 2006; 30: 374–388.
153. Korff S, Harvey BH. Animal models of obsessive-compulsive disorder: rationale to understanding psychobiology and pharmacology. *Psychiatr Clin N Am* 2006; 29: 371–390.
154. Powell SB, Newman HA, Pendergast JF, et al. A rodent model of spontaneous stereotypy: initial characterization of developmental, environmental,

and neurobiological factors. *Physiol Behav* 1999; 66: 355–363.

155. Presti MF, Lewis MH. Striatal opioid peptide content in an animal model of spontaneous stereotypic behavior. *Behav Brain Res* 2005; 157: 363–368.
156. Woods A, Smith C, Szewczak M, et al. Selective serotonin re-uptake inhibitors decrease schedule-induced polydipsia in rats: a potential model for obsessive compulsive disorder. *Psychopharmacology (Berl)* 1993; 112: 195–198.
157. Altemus M, Glowa JR, Galliven E, et al. Effects of serotonergic agents on food-restriction-induced hyperactivity. *Pharmacol Biochem Behav* 1996; 53: 123–131.
158. Broekkamp CL, Jenck F. The relationship between various animal models of anxiety, fear-related psychiatric symptoms and response to serotonergic drugs. In: Bevan P, Cools R, Archer T, eds. *Behavioural Pharmacology of 5-HT.* Hillsdale, NJ: Erlbaum; 1989; 321–335.
159. Broekkamp CL, Rijk HW, Joly-Gelouin D, et al. Major tranquillizers can be distinguished from minor tranquillizers on the basis of effects on marble burying and swim-induced grooming in mice. *Eur J Pharmacol* 1986; 126: 223–229.
160. Gyertyan I. Analysis of the marble burying response: marbles serve to measure digging rather than evoke burying. *Behav Pharmacol* 1995; 6: 24–31.
161. Londei T, Valentini AM, Leone VG. Investigative burying by laboratory mice may involve non-functional, compulsive behaviour. *Behav Brain Res* 1998; 94: 249–254.
162. Njung'e K, Handley SL. Evaluation of marble-burying behavior as a model of anxiety. *Pharmacol Biochem Behav* 1991; 38: 63–67.
163. Szechtman H, Woody E. Obsessive-compulsive disorder as a disturbance of security motivation. *Psychol Rev* 2004; 111: 111–127.
164. Otto MW. Normal and abnormal information processing. A neuropsychological perspective on obsessive compulsive disorder. *Psychiatr Clin N Am* 1992; 15: 825–848.
165. Joel D, Doljansky J. Selective alleviation of compulsive lever-pressing in rats by D1, but not D2, blockade: possible implications for the involvement of D1 receptors in obsessive-compulsive disorder. *Neuropsychopharmacology* 2003; 28: 77–85.
166. Joel D, Ben-Amir E, Doljansky J, et al. 'Compulsive' lever-pressing in rats is attenuated by the serotonin re-uptake inhibitors paroxetine and fluvoxamine but not by the tricyclic antidepressant desipramine or the anxiolytic diazepam. *Behav Pharmacol* 2004; 15: 241–252.
167. Brimberg L, Flaisher-Grinberg S, Schilman EA, et al. Strain differences in 'compulsive' lever-pressing. *Behav Brain Res* 2007; 179: 141–151.
168. Campbell KM, de Lecea L, Severynse DM, et al. OCD-Like behaviors caused by a neuropotentiating transgene targeted to cortical and limbic D1+ neurons. *J Neurosci* 1999; 19: 5044–5053.
169. Nordstrom EJ, Burton FH. A transgenic model of comorbid Tourette's syndrome and obsessive-compulsive disorder circuitry. *Mol Psychiatry* 2002; 7: 617–625, 524.
170. Greer JM, Capecchi MR. Hoxb8 is required for normal grooming behavior in mice. *Neuron* 2002; 33: 23–34.
171. Nonogaki K, Strack AM, Dallman MF, et al. Leptin-independent hyperphagia and type 2 diabetes in mice with a mutated serotonin 5-HT2C receptor gene. *Nat Med* 1998; 4: 1152–1156.
172. Tecott LH, Sun LM, Akana SF, et al. Eating disorder and epilepsy in mice lacking 5-HT2c serotonin receptors. *Nature* 1995; 374: 542–546.
173. Vickers SP, Clifton PG, Dourish CT, et al. Reduced satiating effect of d-fenfluramine in serotonin 5-HT(2C) receptor mutant mice. *Psychopharmacology (Berl)* 1999; 143: 309–314.
174. Chou-Green JM, Holscher TD, Dallman MF, et al. Compulsive behavior in the 5-HT2C receptor knockout mouse. *Physiol Behav* 2003; 78: 641–649.
175. Berridge KC, Aldridge JW, Houchard KR, et al. Sequential super-stereotypy of an instinctive fixed action pattern in hyper-dopaminergic mutant mice: a model of obsessive compulsive disorder and Tourette's. *BMC Biol* 2005; 3: 4.
176. Yadin E, Friedman E, Bridger WH. Spontaneous alternation behavior: an animal model for obsessive-compulsive disorder? *Pharmacol Biochem Behav* 1991; 40: 311–315.
177. Eilam D, Szechtman H. Towards an animal model of obsessive-compulsive disorder (OCD): sensitization to dopamine agonist quinpirole. *Soc Neurosci Abstr* 1995; 21: 192.
178. Szechtman H, Eckert MJ, Tse WS, et al. Compulsive checking behavior of quinpirole-sensitized rats as an animal model of obsessive-compulsive disorder (OCD): form and control. *BMC Neurosci* 2001; 2: 4.
179. Ben-Pazi A, Szechtman H, Eilam D. The morphogenesis of motor rituals in rats treated chronically with the dopamine agonist quinpirole. *Behav Neurosci* 2001; 115: 1301–1317.
180. Zadicario P, Ronen S, Eilam D. Modulation of quinpirole-induced compulsive-like behavior in rats by environmental changes: implications for OCD rituals and for exploration and navigation. *BMC Neurosci* 2007; 8: 23.
181. Meyer V. Modification of expectations in cases with obsessional rituals. *Behav Res Ther* 1966; 4: 273–280.
182. Steketee G. Behavioural assessment and treatment planning with obsessive compulsive disorder: a review emphasizing clinical application. *Behav Ther* 1994; 25: 613–633.
183. Beck AT. *Cognitive Therapy and the Emotional Disorders.* New York, NY: International Universities Press; 1976.
184. Salkovskis PM. Obsessional-compulsive problems: a cognitive-behavioural analysis. *Behav Res Ther* 1985; 23: 571–583.
185. Gava I, Barbui C, Aguglia E, et al. Psychological treatments versus treatment as usual for obsessive compulsive disorder (OCD). *Cochrane Database Syst Rev* 2007:; CD005333.
186. Piccinelli M, Pini S, Bellantuono C, et al. Efficacy of drug treatment in obsessive-compulsive disorder. A meta-analytic review. *Br J Psychiatry* 1995; 166: 424–443.
187. Ellingrod VL. Pharmacotherapy of primary obsessive-compulsive disorder: review of the literature. *Pharmacotherapy* 1998; 18: 936–960.
188. Soomro G, Oakley-Browne M, Doughty C. Serotonin re-uptake inhibitors (SSRIs) versus placebo for

obsessive compulsive disorders (OCD). *Cochrane Database Syst Rev* 2006:; CD001765.

189. March JS, Frances A, Carpenter D, et al. The expert consensus guideline series: treatment of obsessive-compulsive disorder. *J Clin Psychiatry* 1997; 58(suppl 4): 13–72.
190. Delgado PL, Charney DS, Price LH, et al. Serotonin function and the mechanism of antidepressant action. Reversal of antidepressant-induced remission by rapid depletion of plasma tryptophan. *Arch Gen Psychiatry* 1990; 47: 411–418.
191. Aberg-Wistedt A, Hasselmark L, Stain-Malmgren R, et al. Serotonergic 'vulnerability' in affective disorder: a study of the tryptophan depletion test and relationships between peripheral and central serotonin indexes in citalopram-responders. *Acta Psychiatr Scand* 1998; 97: 374–380.
192. Barr LC, Goodman WK, McDougle CJ, et al. Tryptophan depletion in patients with obsessive-compulsive disorder who respond to serotonin reuptake inhibitors. *Arch Gen Psychiatry* 1994; 51: 309–317.
193. Kulz AK, Meinzer S, Kopasz M, et al. Effects of tryptophan depletion on cognitive functioning, obsessive-compulsive symptoms and mood in obsessive-compulsive disorder: preliminary results. *Neuropsychobiology* 2007; 56: 127–131.
194. Smeraldi E, Diaferia G, Erzegovesi S, et al. Tryptophan depletion in obsessive-compulsive patients. *Biol Psychiatry* 1996; 40: 398–402.
195. Huwig-Poppe C, Voderholzer U, Backhaus J, et al. The tryptophan depletion test. Impact on sleep in healthy subjects and patients with obsessive-compulsive disorder. *Adv Exp Med Biol* 1999; 467: 35–42.
196. Berney A, Sookman D, Leyton M, et al. Lack of effects on core obsessive-compulsive symptoms of tryptophan depletion during symptom provocation in remitted obsessive-compulsive disorder patients. *Biol Psychiatry* 2006; 59: 853–857.
197. Marks IM, Stern RS, Mawson D, et al. Clomipramine and exposure for obsessive-compulsive rituals: i. *Br J Psychiatry* 1980; 136: 1–25.
198. Marks IM, Lelliott P, Basoglu M, et al. Clomipramine, self-exposure and therapist-aided exposure for obsessive-compulsive rituals. *Br J Psychiatry* 1988; 152: 522–534.
199. Cottraux J, Mollard E, Bouvard M, et al. A controlled study of fluvoxamine and exposure in obsessive-compulsive disorder. *Int Clin Psychopharmacol* 1990; 5: 17–30.
200. Hohagen F, Winkelmann G, Rasche-Ruchle H, et al. Combination of behaviour therapy with fluvoxamine in comparison with behaviour therapy and placebo. Results of a multicentre study. *Br J Psychiatry Suppl* 1998;; 71–78.
201. van Balkom AJ, de Haan E, van Oppen P, et al. Cognitive and behavioral therapies alone versus in combination with fluvoxamine in the treatment of obsessive compulsive disorder. *J Nerv Ment Dis* 1998; 186: 492–499.
202. Foa EB, Liebowitz MR, Kozak MJ, et al. Randomized, placebo-controlled trial of exposure and ritual prevention, clomipramine, and their combination in the treatment of obsessive-compulsive disorder. *Am J Psychiatry* 2005; 162: 151–161.
203. Pediatric OCD Treatment Study Team. Cognitive-behavior therapy, sertraline, and their combination for children and adolescents with obsessive-compulsive disorder: the Pediatric OCD Treatment Study (POTS) randomized controlled trial. *JAMA* 2004; 292: 1969–1976.
204. Tenneij NH, van Megen HJ, Denys DA, et al. Behavior therapy augments response of patients with obsessive-compulsive disorder responding to drug treatment. *J Clin Psychiatry* 2005; 66: 1169–1175.
205. Simpson HB, Liebowitz MR, Foa EB, et al. Post-treatment effects of exposure therapy and clomipramine in obsessive-compulsive disorder. *Depress Anxiety* 2004; 19: 225–233.
206. Baxter LR Jr, Schwartz JM, Bergman KS, et al. Caudate glucose metabolic rate changes with both drug and behavior therapy for obsessive-compulsive disorder. *Arch Gen Psychiatry* 1992; 49: 681–689.
207. Schwartz JM, Stoessel PW, Baxter LR Jr, et al. Systematic changes in cerebral glucose metabolic rate after successful behavior modification treatment of obsessive-compulsive disorder. *Arch Gen Psychiatry* 1996; 53: 109–113.
208. Goodman WK, Price LH, Rasmussen SA, et al. Efficacy of fluvoxamine in obsessive-compulsive disorder: a double-blind comparison with placebo. *Arch Gen Psychiatry* 1989; 46: 36–44.
209. McDonough M, Kennedy N. Pharmacological management of obsessive-compulsive disorder: a review for clinicians. *Harvard Rev Psychiatry* 2002; 10: 127–137.
210. Hollander E, Bienstock CA, Koran LM, et al. Refractory obsessive-compulsive disorder: state-of-the-art treatment. *J Clin Psychiatry* 2002; 63(suppl 6): 20–29.
211. Goodman WK, McDougle CJ, Barr LC, et al. Biological approaches to treatment-resistant obsessive compulsive disorder. *J Clin Psychiatry* 1993; 54(suppl 6): 16–26.
212. Pallanti S, Hollander E, Bienstock C, et al. Treatment non-response in OCD: methodological issues and operational definitions. *Int J Neuropsychopharmacol* 2002; 5: 181–191.
213. Kaplan A, Hollander E. A review of pharmacologic treatments for obsessive-compulsive disorder. *Psychiatr Serv* 2003; 54: 1111–1118.
214. Alacron RD, Libb JW, Spitler D. A predictive study of obsessive compulsive response to clomipramine. *J Clin Psychopharmacol* 1993; 13: 210–213.
215. Erzegovesi S, Cavellini MC, Cavedini P, et al. Clinical predictors of drug response in obsessive compulsive disorder. *J Clin Psychopharmacol* 2001; 21: 272–275.
216. Ravizza L, Barzega G, Bellino S, et al. Predictors of drug treatment response in obsessive-compulsive disorder. *J Clin Psychiatry* 1995; 56: 368–373.
217. Wheadon DE, Bushnell WD, Steiner M. A fixed-dose comparison of 20, 40 or 60 mg paroxetine or placebo in the treatment of obsessive-compulsive disorder. Paper presented at: 32nd Annual Meeting of the American College of Neuropsychopharmacology; December 13–17, 1993; Honolulu, Hawaii.
218. Goodman W, Price L, Rasmussen S, et al. The Yale-Brown Obsessive Compulsive Scale (Y-BOCS), part I: development, use and reliability. *Arch Gen Psychiatry* 1989; 46: 1006–1011.

219. Goodman W, Price L, Rasmussen S, et al. The Yale-Brown Obsessive Compulsive Scale (Y-BOCS), part II: validity. *Arch Gen Psychiatry* 1989; 46: 1012–1016.
220. Diering SL, Bell WO. Functional neurosurgery for psychiatric disorders: a historical perspective. *Stereotact Funct Neurosurg* 1991; 57: 175–194.
221. Mashour GA, Walker EE, Martuza RL. Psychosurgery: past, present, and future. *Brain Res Rev* 2005; 48: 409–419.
222. Hollander E, Stein DJ, Kwon JH, et al. Psychosocial function and economic costs of obsessive-compulsive disorder. *CNS Spectr* 1997; 2: 16–25.
223. Koran LM. Quality of life in obsessive-compulsive disorder. *Psychiatr Clin N Am* 2000; 23: 509–517.
224. Fineberg NA, Gale TM. Evidence-based pharmacotherapy of obsessive compulsive disorder. *Int J Neuropsychopharmacol* 2004; 7: 1–23.
225. Jacobsen FM. Risperidone in the treatment of affective illness and obsessive-compulsive disorder. *J Clin Psychiatry* 1995; 56: 423–429.
226. McDougle CJ, Goodman WK, Price LH, et al. Neuroleptic addition in fluvoxamine-refractory obsessive compulsive disorder. *Am J Psychiatry* 1990; 147: 652–654.
227. McDougle CJ, Fleishmann RL, Epperson CN, et al. Risperidone addition in fluvoxamine-refractory obsessive-compulsive disorder: three cases. *J Clin Psychiatry* 1995; 56(suppl 11): 526–528.
228. Stein DJ, Bouwer C, Hawkridge S, Emsley RA. Risperidone augmentation of serotonin reuptake inhibitors in obsessive compulsive disorder. *J Clin Psychiatry* 1997; 58(suppl 3): 119–122.
229. McDougle CJ, Price LH, Goodman WK, et al. A controlled trial of lithium augmentation in fluvoxamine refractory obsessive compulsive disorder: lack of efficacy. *J Clin Psychopharmacol* 1991; 11: 175–184.
230. Pigott TA, L'Heureux F, Hill JL, et al. A double-blind study of adjuvant buspirone hydrochloride in clomipramine-treated patients with obsessive-compulsive disorder. *J Clin Psychopharmacol* 1992; 12: 11–18.
231. McDougle CJ, Goodman WK, Leckman JF, et al. Haloperidol addition in fluvoxamine-refractory obsessive-compulsive disorder: a double-blind, placebo-controlled study in patients with and without tics. *Arch Gen Psychiatry* 1994; 51: 302–308.
232. Li X, May RS, Tolbert LC, et al. Risperidone and haloperidol augmentation of serotonin reuptake inhibitors in refractory obsessive compulsive disorder: a crossover study. *J Clin Psychiatry* 2005; 66: 736–743.
233. McDougle CJ, Epperson CN, Pelton GH, et al. A double-blind, placebo-controlled study of risperidone addition in serotonin reuptake inhibitor refractory obsessive-compulsive disorder. *Arch Gen Psychiatry* 2000; 57: 794–801.
234. Hollander E, Baldini RN, Sood E, Pallanti S. Risperidone augmentation in treatment-resistant obsessive compulsive disorder: a double-blind, placebo-controlled study. *Int J Neuropsychopharmacol* 2003; 6: 397–401.
235. Erzegovesi S, Guglielmo E, Siliprandi F, Bellodi L. Low dose risperidone augmentation of fluvoxamine treatment in obsessive-compulsive disorder: a double-blind, placebo-controlled study. *Eur Neuropsychopharmacol* 2005; 15: 69–74.
236. Bystritsky A, Ackerman DL, Rosen RM, et al. Augmentation of serotonin reuptake inhibitors in refractory obsessive compulsive disorder using adjunctive olanzapine: a placebo-controlled trial. *J Clin Psychiatry* 2004; 65: 565–568.
237. Shapira NA, Ward HE, Mandoki M, et al. A double-blind, placebo controlled trial of olanzapine addition in fluoxetine refractory obsessive-compulsive disorder. *Biol Psychiatry* 2004; 55: 553–555.
238. Denys D, de Geus F, van Megen HJ, Westenberg HG. A double-blind, randomized, placebo-controlled trial of quetiapine addition in patients with obsessive compulsive disorder refractory to serotonin reuptake inhibitors. *J Clin Psychiatry* 2004; 65: 1040–1048.
239. Carey PD, Vythilingum B, Seedat S, et al. Quetiapine augmentation of SRIs in treatment refractory obsessive-compulsive disorder: a double-blind, randomised, placebo-controlled study [ISRCTN83050762]. *BMC Psychiatry* 2005; 5: 5.
240. Fineberg NA, Sivakumaran T, Roberts A, Gale T. Adding quetiapine to SRI in treatment-resistant obsessive compulsive disorder: a randomized controlled treatment study. *Int Clin Psychopharmacol* 2005; 20: 223–226.
241. Atmaca M, Kuloglu M, Tezcan E, Gecici O. Quetiapine augmentation in patients with treatment resistant obsessive-compulsive disorder: a single-blind, placebo-controlled study. *Int Clin Psychopharmacol* 2002; 17: 115–119.
242. Bloch MH, Landeros-Weisenberger A, Kelmendi B, et al. A systematic review: antipsychotic augmentation with treatment refractory obsessive-compulsive disorder. *Mol Psychiatry* 2006; 11: 622–632.
243. Maina G, Albert U, Ziero S, Bogetto F. Antipsychotic augmentation for treatment resistant obsessive-compulsive disorder: what if antipsychotic is discontinued? *Int Clin Psychopharmacol* 2003; 18: 23–28.
244. Joffe RT, Swinson RP. Carbamazepine in obsessive compulsive disorder. *Biol Psychiatry* 1987; 22: 1169–1171.
245. Cora-Locatelli G, Greenberg BD, Martin J, Murphy DL. Gabapentin augmentation for fluoxetine-treated patients with obsessive-compulsive disorder. *J Clin Psychiatry* 1998; 59: 480–481.
246. Sporn J, Smith M, Jersino JM, et al. A double-blind, placebo-controlled trial of gabapentin (GBP) augmentation of fluoxetine for treatment of obsessive-compulsive disorder (OCD). Poster presented at: New Clinical Drug Evaluation Unit Program; May 28–31, 2001; Phoenix, AZ.
247. van Ameringen M, Mancini C, Patterson B, Bennett M. Topiramate augmentation in treatment-resistant obsessive-compulsive disorder: a retrospective, open label case series. *Depress Anxiety* 2006; 23: 1–5.
248. George MS. New methods of brain stimulation are improving research and therapy in obsessive-compulsive disorder. *CNS Spectr* 2000; 5(suppl 4): 12–17.
249. Jenike MA, Rauch SL, Baer L, Rasmussen SA. Neurosurgical treatment of obsessive-compulsive disorder. In: Jenike MA, Baer L, Minichiello WE, eds. Obsessive-Compulsive Disorders: Practical Management. St Louis, MO: Mosby; 1998; 592–610.

250. Freeman C, Crossley D, Eccleston D. Neurosurgery for Mental Disorder: Report from the Neurosurgery Working Group of the Royal College of Psychiatrists. London, United Kingdom: Royal College of Psychiatrists; 2000; CR89.
251. Greenberg BD, Price LH, Rauch SL, et al. Neurosurgery for intractable obsessive-compulsive disorder and depression: critical issues. *Neurosurg Clin N Am* 2003; 14: 199–212.
252. Lisanby SH, ed. Brain Stimulation in Psychiatric Treatment. Arlington, VA: American Psychiatric Publishing; 2004.
253. American Psychiatric Association. The Practice of Electroconvulsive Therapy: Recommendations for Treatment, Training and Privileging. A Task Force Report of the American Psychiatric Association. 2nd ed. Washington, DC: American Psychiatric Association; 2001.
254. Maletzky B, McFarland B, Burt A. Refractory obsessive compulsive disorder and ECT. *Convuls Ther* 1994; 10: 34–42.
255. Khanna S, Gangadhar BN, Sinha V, et al. Electroconvulsive therapy in obsessive-compulsive disorder. *Convuls Ther* 1988; 4: 314–320.
256. Mellman LA, Gorman JM. Successful treatment of obsessive-compulsive disorder with ECT. *Am J Psychiatry* 1984; 141: 596–597.
257. Husain MM, Lewis SF, Thornton WL. Maintenance ECT for refractory obsessive-compulsive disorder. *Am J Psychiatry* 1993; 150: 1899–1900.
258. Casey DA, Davis MH. Obsessive-compulsive disorder responsive to electroconvulsive therapy in an elderly woman. *South Med J* 1994; 87: 862–864.
259. Lavin MR, Halligan P. ECT for comorbid obsessive compulsive disorder and schizophrenia. *Am J Psychiatry* 1996; 153: 1652–1653.
260. Thomas SG, Kellner CH. Remission of major depression and obsessive-compulsive disorder after a single unilateral ECT. *J ECT* 2003; 19: 50–51.
261. Strassnig M, Riedel M, Muller N. Electroconvulsive therapy in a patient with Tourette's syndrome and comorbid obsessive compulsive disorder. *World J Biol Psychiatry* 2004; 5: 164–166.
262. Beale MD, Kellner CH, Pritchett JT, et al. ECT for OCD. *J Clin Psychiatry* 1995; 56: 81–82.
263. Chung MS, Huang SC, Tsai SJ, et al. Mania associated with electroconvulsive therapy in a patient with obsessive compulsive disorder. *Int Med J* 2001; 8: 185–186.
264. Rapoport M, Feder V, Sandor P. Response of major depression and Tourette's syndrome to ECT: a case report. *Psychosom Med* 1998; 60: 528–529.
265. Husain SS, Kevan IM, Linnell R, et al. Electroconvulsive therapy in depressive illness that has not responded to drug treatment. *J Affect Disord* 2004; 83: 121–126.
266. Tharyan P, Adams CE. Electroconvulsive therapy for schizophrenia. *Cochrane Database Syst Rev* 2005; CD000076.
267. National Institute for Health and Clinical Excellence. *Guidance on the Use of Electroconvulsive Therapy. Technology Appraisal* 59. London, United Kingdom: National Institute for Health and Clinical Excellence; 2003.
268. Sackeim HA. Magnetic stimulation therapy and ECT. *Convuls Ther* 1994; 10: 255–258.
269. Lisanby SH. Update on magnetic seizure therapy (MST): a novel form of convulsive therapy. *J ECT* 2002; 18: 182–188.
270. Dell'Osso B, Altamura AC, Allen A, Hollander E. Brain stimulation techniques in the treatment of obsessive-compulsive disorder: current and future directions. *CNS Spectr* 2005; 10: 966–979, 983.
271. Barker AT, Jalinous R, Freeston IL. Non-invasive magnetic stimulation of human motor cortex. *Lancet* 1985; 1: 1106–1107.
272. Bickford RG, Guidi M, Fortesque P, et al. Magnetic stimulation of human peripheral nerve and brain: response enhancement by combined magnetoelectrical technique. *Neurosurgery* 1987; 20: 110–116.
273. George MS, Lisanby SH, Sackeim HA, et al. Transcranial magnetic stimulation: applications in neuropsychiatry. *Arch Gen Psychiatry* 1999; 56: 300–311.
274. Pascual-Leone A, Valls-Sole J, Wasserman EM, et al. Responses to rapid rate transcranial magnetic stimulation of the human motor cortex. *Brain* 1994; 117: 847–858.
275. Pascual-Leone A, Catala MD, Pascual-Leone PA. Lateralized effect of rapid-rate transcranial magnetic stimulation of the prefrontal cortex on mood. *Neurology* 1996; 46: 499–502.
276. Chen R, Classen J, Gerloff C, et al. Depression of motor cortex excitability by low-frequency transcranial magnetic stimulation. *Neurology* 1997; 48: 1398–1403.
277. Bohning DE, Shastry A, Wassermann EM, et al. BOLD-fMRI response to single-pulse transcranial magnetic stimulation (TMS). *J Magn Reson Imaging* 2000; 11(569): 574.
278. Catafau A, Perez V, Gironell A, et al. SPECT mapping of cerebral activity changes induced by repetitive transcranial magnetic stimulation in depressed patients: a pilot study. *Psychiatry Res* 2001; 106: 151–160.
279. Schaepfer TE, Kosel M. Transcranial magnetic stimulation in depression. In: Lisanby SH, ed. *Brain Stimulation in Psychiatric Treatment*. Arlington, VA: American Psychiatric Publishing; 2004; 1–22.
280. Martin JL, Barbanoj MJ, Perez V, et al. Transcranial magnetic stimulation for the treatment of obsessive-compulsive disorder. *Cochrane Database Syst Rev* 2003; CD003387.
281. Bohning DE, Pecheny AP, Epstein CM, et al. Mapping transcranial magnetic stimulation (TMS) in vivo with MRI. *Neuroreport* 1997; 8: 2535–2538.
282. Paus T, Jech R, Thompson J, et al. Transcranial magnetic stimulation during positron emission tomography: a new method for studying connectivity of the human cerebral cortex. *J Neurosci* 1997; 17: 3178–3184.
283. George MS, Nahas Z, Kozel A, et al. Mechanisms and the current state of transcranial magnetic stimulation. *CNS Spectr* 2003; 496–502, 511–514.
284. Roth Y, Zangen A, Hallett M. A coil design for transcranial magnetic stimulation of deep brain regions. *J Clin Neurophysiol* 2002; 19: 361–370.
285. Ferrao YA, Shavitt RG, Bedin NR, et al. Clinical features associated to refractory obsessive-compulsive disorder. *J Affect Disord* 2006; 94: 199–209.

286. Lopes AC, de Mathis ME, Canteras MM, et al. Update on neurosurgical treatment for obsessive compulsive disorder. *Rev Bras Psiquiatr* 2004; 26: 62–66.
287. Schruers K, Koning K, Luermans J, et al. Obsessive-compulsive disorder: a critical review of therapeutic perspectives. *Acta Psychiatr Scand* 2005; 111: 261–271.
288. Kiloh LG, Smith JS, Johnston GF. Psychosurgery (limbic surgery). In: Kiloh LG, Smith JS, Johnston GF, eds. *Physical Treatments in Psychiatry*. Melbourne, Australia: Blackwell; 1988; 277–333.
289. Waziri, R. Psychosurgery for anxiety and obsessive-compulsive disorders. In: Noyes R, Roth M, Burrows GD, eds. *The Treatment of Anxiety: Handbook of Anxiety*. Amsterdam, The Netherlands: Elsevier; 1990.
290. Kringelbach ML, Jenkinson N, Owen SL, Aziz TZ. Translational principles of deep brain stimulation. *Nat Rev Neurosci* 2007; 8: 623–635.
291. Rezai AR, Mogilner AY, Cappell J, et al. Integration of functional brain mapping in image-guided neurosurgery. *Acta Neurochir Suppl* 1997; 68: 85–89.
292. Schuurman PR, Bosch DA, Bossuyt PM, et al. A comparison of continuous thalamic stimulation and thalamotomy for suppression of severe tremor. *N Engl J Med* 2000; 342: 461–468.
293. Schruers K, Koning K, Luermans J, et al. Obsessive-compulsive disorder: a critical review of therapeutic perspectives. *Acta Psychiatr Scand* 2005; 111: 261–271.
294. Lozano AM, Dostrovsky J, Chen R, Ashby P. Deep brain stimulation for Parkinson's disease: disrupting the disruption. *Lancet Neurol* 2002; 1: 225–231.
295. McIntyre CC, Savasta M, Walter BL, Vitek JL. How does deep brain stimulation work? Present understanding and future questions. *J Clin Neurophysiol* 2004; 21: 40–50.
296. Dostrovsky JO, Lozano AM. Mechanisms of deep brain stimulation. *Mov Disord* 2002; 17(suppl 3): S63–S68.
297. Benabid, AL, Pollak P, Gervason C, et al. Long term suppression of tremor by chronic stimulation of ventral intermediate thalamic nucleus. *Lancet* 1991; 337: 402–406.
298. Blond S, Caparros-Lefebvre D, Parker F, et al. Control of tremor and involuntary movement disorders by chronic stereotactic stimulation of the ventral intermediate thalamic nucleus. *J Neurosurg* 1992; 77: 62–68.
299. Tass PA, Klosterkotter J, Schneider F, et al. Obsessive-compulsive disorder: development of demand-controlled deep brain stimulation with methods from stochastic phase resetting. *Neuropsychopharmacology* 2003; 28(suppl 1): S27–S34.
300. Greenberg BD. Deep brain stimulation in psychiatry. In: Lisanby SH, ed. *Brain Stimulation in Psychiatric Treatment*. Arlington, VA: American Psychiatric Publishing; 2004; 53–65.
301. Boraud T, Bezard E, Bioulac B, Gross C. High frequency stimulation of the internal globus pallidus (GPi) simultaneously improves Parkinsonian symptoms and reduces the firing frequency of GPi neurons in the MPTP-treated monkey. *Neurosci Lett* 1996; 215: 17–20.
302. Benazzouz A, Hallett M. Mechanism of action of deep brain stimulation. *Neurology* 2000; 55(suppl 6): S13–S16.
303. Hashimoto T, Elder CM, Okun MS, et al. Stimulation of the subthalamic nucleus changes the firing pattern of pallidal neurons. *J Neurosci* 2003; 23: 1916–1923.
304. Devos S, Labyt E, Derambure P, et al. Subthalamic nucleus stimulation modulates motor cortex oscillatory activity in Parkinson's disease. *Brain* 2004; 127: 408–419.
305. Hamani C, Lozano AM. Physiology and pathophysiology of Parkinson's disease. *Ann N Y Acad Sci* 2003; 991: 15–21.
306. Hamani C, Richter E, Schwalb JM, Lozano AM. Bilateral subthalamic nucleus stimulation for Parkinson's disease: a systematic review of the clinical literature. *Neurosurgery* 2005; 56: 1313–1324.
307. Hamani C, Neimat J, Lozano AM. Deep brain stimulation for the treatment of Parkinson's disease. *J Neural Transm Suppl* 2006; 70: 393–399.
308. Lozano AM, Mahant N. Deep brain stimulation surgery for Parkinson's disease: mechanisms and consequences. *Parkinsonism Relat Disord* 2004; 10(suppl 1): S49–S57.
309. Deuschl G, Bain P. Deep brain stimulation for tremor [correction of trauma]: patient selection and evaluation. *Mov Disord* 2002; 17(S3): 102–111.
310. Tasker RR. Deep brain stimulation is preferable to thalamotomy for tremor suppression. *Surg Neurol* 1998; 49: 145–153.
311. Pollak P, Fraix V, Krack P, et al. Treatment results: Parkinson's disease. *Mov Disord* 2002; 17(suppl 3): 75–83.
312. Rauch SL, Dougherty DD, Malone D, et al. A functional neuroimaging investigation of deep brain stimulation in patients with obsessive-compulsive disorder. *J Neurosurg* 2006; 104: 558–565.
313. Nuttin B, Cosyns P, Demeulemeester H, et al. Electrical stimulation in anterior limbs of internal capsules in patients with obsessive-compulsive disorder. *Lancet* 1999; 354: 1526.
314. Nuttin BJ, Gabriels L, van Kuyck K, et al. Electrical stimulation of the anterior limbs of the internal capsules in patients with severe obsessive compulsive disorder: anecdotal reports. *Neurosurg Clin N Am* 2003; 14: 267–274.
315. Nuttin BJ, Gabriels LA, Cosyns PR, et al. Long-term electrical capsular stimulation in patients with obsessive-compulsive disorder. *Neurosurgery* 2003; 52: 1263–1272.
316. Gabriels L, Cosyns P, Nuttin B, et al. Deep brain stimulation for treatment-refractory obsessive-compulsive disorder: psychopathological and neuropsychological outcome in three cases. *Acta Psychiatr Scand* 2003; 107: 275–282.
317. de la Fuente-Fernández R. Uncovering the hidden placebo effect in deep-brain stimulation for Parkinson's disease. *Parkinsonism Relat Disord* 2004; 10: 125–127.
318. Sturm V, Lenartz D, Koulousakis A, et al. The nucleus accumbens: a target for deep brain stimulation in obsessive-compulsive- and anxiety-disorders. *J Chem Neuroanat* 2003; 26: 293–299.
319. Shumyatsky GP, Tsvetkov E, Malleret G, et al. Identification of a signaling network in lateral nucleus of amygdala important for inhibiting memory

specifically related to learned fear. *Cell* 2002; 111: 905–918.
320. Grace AA. Cortical regulation of subcortical dopamine systems and its possible relevance to schizophrenia. *J Neural Trans Gen* 1993; 9: 111–134.
321. Mulder AB, Hodenpijl MG, Lopes da Silva FH. Electrophysiology of the hippocampal and amygdaloid projections to the nucleus accumbens of the rat: convergence, segregation, and interaction of inputs. *J Neurosci* 1998; 18: 5059–5102.
322. Rosenkranz JA, Grace AA. Modulation of basolateral amygdale neuronal firing and afferent drive by dopamine receptor activation in vivo. *J Neurosci* 1999; 15: 11027–11039.
323. Abelson JL, Curtis GC, Sagher O, et al. Deep brain stimulation for refractory obsessive-compulsive disorder. *Biol Psychiatry* 2005; 57: 510–516.
324. Greenberg BD, Malone DA, Friehs GM, et al. Three-year outcomes in deep brain stimulation for highly resistant obsessive-compulsive disorder. *Neuropsychopharmacology* 2006; 31: 2384–2393.
325. Anderson D, Ahmed A. Treatment of patients with intractable obsessive-compulsive disorder with anterior capsular stimulation. Case report. *J Neurosurg* 2003; 98: 1104–1108.
326. Fontaine D, Mattei V, Borg M, et al. Effect of subthalamic nucleus stimulation on obsessive-compulsive disorder in a patient with Parkinson disease. Case report. *J Neurosurg* 2004; 100: 1084–1086.
327. Mallet L, Mesnage V, Houeto JL, et al. Compulsions, Parkinson's disease, and stimulation. *Lancet* 2002; 360: 1302–1304.
328. Aouizerate B, Cuny E, Martin-Guehl C, et al. Deep brain stimulation of the ventral caudate nucleus in the treatment of obsessive-compulsive disorder and major depression. Case report. *J Neurosurg* 2004; 101: 682–686.
329. Mayberg HS, Lozano AM, Voon V, et al. Deep brain stimulation for treatment-resistant depression. *Neuron* 2005; 45: 651–660.
330. Greenberg BD, Nahas Z, Carpenter LL. Current status of deep brain stimulation. *Prim Psychiatry* 2005; 12: 59–64.
331. Alterman RL, Miravite J, Weisz D, et al. Sixty hertz pallidal deep brain stimulation for primary torsion dystonia. *Neurology* 2007; 69: 681–688.
332. Koran LM, Quirk T, Lorberbaum JP, Elliott M. Mirtazapine treatment of obsessive-compulsive disorder. *J Clin Psychopharmacol* 2001; 21: 537–539.
333. Koran LM, Gamel NN, Choung HW, et al. Mirtazapine for obsessive-compulsive disorder: an open trial followed by double-blind discontinuation. *J Clin Psychiatry* 2005; 66: 515–520.
334. Jenike MA, Baer L, Minichiello WE, et al. Placebo-controlled trial of fluoxetine and phenelzine for obsessive-compulsive disorder. *Am J Psychiatry* 1997; 154: 1261–1264.
335. Mundo E, Guglielmo E, Bellodi L. Effect of adjuvant pindolol on the antiobsessional response to fluvoxamine: a double-blind, placebo-controlled study. *Int Clin Psychopharmacol* 1998; 13: 219–224.
336. Koran LM, Mueller K, Maloney A. Will pindolol augment the response to a serotonin reuptake inhibitor in obsessive-compulsive disorder? *J Clin Psychopharmacol* 1996; 16: 253–254.
337. Hewlett WA, Schmid SP, Salomon RM. Pilot trial of ondansetron in the treatment of 8 patients with obsessive-compulsive disorder. *J Clin Psychiatry* 2003; 64: 1025–1030.
338. Levine J. Controlled trials of inositol in psychiatry. *Eur Neuropsychopharmacol* 1997; 7: 147–155.
339. Fux M, Benjamin J, Belmaker RH. Inositol versus placebo augmentation of serotonin reuptake inhibitors in the treatment of obsessive-compulsive disorder: a double-blind cross-over study. *Int J Neuropsychopharmacol* 1999; 2: 193–195.
340. Insel TR, Hamilton JA, Guttmacher LB, Murphy DL. D-amphetamine in obsessive-compulsive disorder. *Psychopharmacology (Berl)* 1983; 80: 231–235.
341. Shapira NA, Keck PE Jr, Goldsmith TD, et al. Open-label pilot study of tramadol hydrochloride in treatment-refractory obsessive-compulsive disorder. *Depress Anxiety* 1997; 6: 170–173.
342. Rauch S. Neuroimaging research and the neurobiology of obsessive-compulsive disorder: where do we go from here? *Biol Psychiatry* 2000; 47: 168–170.
343. Finn DA, Rutledge-Gorman MT, Crabbe JC. Genetic animal models of anxiety. *Neurogenetics* 2003; 4: 109–135.
344. Grados M, Wilcox HC. Genetics of obsessive-compulsive disorder: a research update. *Expert Rev Neurother* 2007; 7: 967–980.
345. Leckman JF, Zhang H, Alsobrook JP, Pauls DL. Symptom dimensions in obsessive-compulsive disorder: toward quantitative phenotypes. *Am J Med Genet* 2001; 105: 28–30.
346. Neale BM, Sham PC. The future of association studies: gene-based analysis and replication. *Am J Hum Genet* 2004; 75: 353–362.
347. Tsaltas E, Kontis D, Chrysikakou S, et al. Reinforced spatial alternation as an animal model of obsessive-compulsive disorder (OCD): investigation of 5-HT2D and 5-HT1D receptor involvement in OCD pathophysiology. *Biol Psychiatry* 2005; 57: 1176–1185.
348. Joel D, Avisar A. Excessive lever pressing following post-training signal attenuation in rats: a possible animal model of obsessive compulsive disorder? *Behav Brain Res* 2001; 123: 77–87.
349. Savage CR, Keuthen NJ, Jenike MA. Recall and recognition memory in obsessive-compulsive disorder. *J Neuropsychiatry Clin Neurosci* 1996; 8: 99–103.
350. Stein DJ. Neurobiology of the obsessive-compulsive spectrum disorders. *Biol Psychiatry* 2000; 47: 296–304.
351. Greenberg BD, George MS, Martin JD, et al. Effect of prefrontal repetitive transcranial magnetic stimulation in obsessive-compulsive disorder: a preliminary study. *Am J Psychiatry* 1997; 154: 867–869.
352. Alonso P, Pujol J, Cardoner N, et al. Right prefrontal repetitive transcranial magnetic stimulation in obsessive-compulsive disorder: a double-blind placebo-controlled study. *Am J Psychiatry* 2001; 158: 1143–1145.
353. Sachdev PS, McBride R, Loo CK, et al. Right versus left prefrontal transcranial magnetic stimulation for obsessive-compulsive disorder: a preliminary investigation. *J Clin Psychiatry* 2001; 62: 981–984.
354. Mantovani A, Lisanby SH, Pieraccini F, et al. Repetitive transcranial magnetic stimulation (rTMS)

in the treatment of obsessive-compulsive disorder (OCD) and Tourette's syndrome (TS). *Int J Neuropsychopharmacol* 2006; 9: 95–100.
355. Strom-Olsen R, Carlisle S. Bi-frontal stereotactic tractotomy: a follow-up study of its effects on 210 patients. *Br J Psychiatry* 1971; 118: 141–154.
356. Goktepe EO, Young LB, Bridges PK. A further review of the results of stereotactic subcaudate tractotomy. *Br J Psychiatry* 1975; 126: 270–280.
357. Woerdeman PA, Willems PW, Noordmans HJ, et al. Frameless stereotactic subcaudate tractotomy for intractable obsessive-compulsive disorder. *Acta Neurochir (Wien)* 2006; 148: 633–637.
358. Leksell L. A stereotaxic apparatus for intracerebral surgery. *Acta Chir Scand* 1949; 99: 229–233.
359. Herner T. Treatment of mental disorders with frontal stereotactic thermo-lesions: a follow-up of 116 cases. *Acta Psychiatr Scand Suppl* 1961; 36.
360. Bingley T, Leksell L, Meyerson BA, et al. Long term results of stereotactic capsulotomy in chronic obsessive-compulsive neurosis. In: Sweet WH, Borador Alcalde S, Martin-Rodriguez JG, eds. *Neurosurgical Treatment in Psychiatry, Pain, and Epilepsy:* Proceedings of the Fourth World Congress of Psychiatric Surgery, September 7–10, 1975, Madrid, Spain. Baltimore, MD: University Park Press; 1977; 287.
361. Fodstad H, Strandman E, Karlsson B, et al. Treatment of chronic obsessive compulsive states with stereotactic anterior capsulotomy or cingulotomy. *Acta Neurochir (Wien)* 1982; 62: 1–23.
362. Guo WY, Lindquist C, Kihlström P, et al. *Radionecrosis Created in the Internal Capsule for Psychosurgery with the Gamma Knife [dissertation].* Stockholm, Sweden: Karolinska Institut; 1993.
363. Mindus P, Rasmussen SA, Lindquist C. Neurosurgical treatment for refractory obsessive-compulsive disorder: implications for understanding frontal lobe function. *J Neuropsychiatry Clin Neurosci* 1994; 6: 467–477.
364. Lippitz BE, Mindus P, Meyerson BA, et al. Lesion topography and outcome after thermocapsulotomy or gamma knife capsulotomy for obsessive-compulsive disorder: relevance of the right hemisphere. *Neurosurgery* 1999; 44: 452–460.
365. Christensen DD, Laitinen LV, Schmidt LJ, et al. Anterior capsulotomy for treatment of refractory obsessive-compulsive disorder: results in a young and an old patient. *Stereotact Funct Neurosurg* 2002; 79: 234–244.
366. Oliver B, Gascon J, Aparicio A, et al. Bilateral anterior capsulotomy for refractory obsessive-compulsive disorders. *Stereotact Funct Neurosurg* 2003; 81: 90–95.
367. Tippin J, Henn FA. Modified leukotomy in the treatment of intractable obsessional neurosis. *Am J Psychiatry* 1982; 139: 1601–1603.
368. Jenike MA, Baer L, Ballantine T, et al. Cingulotomy for refractory obsessive-compulsive disorder. A long-term follow-up of 33 patients. *Arch Gen Psychiatry* 1991; 48: 548–555.
369. Baer L, Rauch SL, Ballantine HT Jr, et al. Cingulotomy for intractable obsessive-compulsive disorder. Prospective long-term follow-up of 18 patients. *Arch Gen Psychiatry* 1995; 52: 384–392.
370. Dougherty DD, Baer L, Cosgrove GR, et al. Prospective long-term follow-up of 44 patients who received cingulotomy for treatment-refractory obsessive-compulsive disorder. *Am J Psychiatry* 2002; 159: 269–275.
371. Spangler WJ, Cosgrove GR, Ballantine HT, et al. Magnetic resonance image-guided stereotactic cingulotomy for intractable psychiatric disease. *Neurosurgery* 1996; 38: 1071–1076.
372. Kim CH, Chang JW, Koo MS, et al. Anterior cingulotomy for refractory obsessive-compulsive disorder. *Acta Psychiatr Scand* 2003; 107: 283–290.
373. Richter EO, Davis KD, Hamani C, et al. Cingulotomy for psychiatric disease: microelectrode guidance, a callosal reference system for documenting lesion location, and clinical results. *Neurosurgery* 2004; 54: 622–628.
374. Jung HH, Kim CH, Chang JH, et al. Bilateral anterior cingulotomy for refractory obsessive-compulsive disorder: long-term follow-up results. *Stereotact Funct Neurosurg* 2006; 84: 184–189.
375. Cumming S, Hay P, Lee T, et al. Neuropsychological outcome from psychosurgery for obsessive-compulsive disorder. *Aust N Z J Psychiatry* 1995; 29: 293–298.
376. Hay P, Sachdev P, Cumming S, et al. Treatment of obsessive-compulsive disorder by psychosurgery. *Acta Psychiatr Scand* 1993; 87: 197–207.
377. Sachdev P, Hay P. Does neurosurgery for obsessive-compulsive disorder produce personality change? *J Nerv Ment Dis* 1995; 183: 408–413.
378. Sachdev P. Site and size of lesion and psychosurgical outcome in obsessive compulsive disorder: a magnetic resonance imaging study. *Biol Psychiatry* 1996; 39: 739–742.
379. Irle E, Exner C, Thielen K, et al. Obsessive-compulsive disorder and ventromedial frontal lesions: clinical and neuropsychological findings. *Am J Psychiatry* 1998; 155: 255–263.
380. Sachdev P, Trollor J, Walker A, et al. Bilateral orbitomedial leucotomy for obsessive-compulsive disorder: a single-case study using positron emission tomography. *Aust N Z J Psychiatry* 2001; 35: 684–690.
381. Montoya A, Weiss AP, Price BH, et al. Magnetic resonance imaging guided stereotactic limbic leukotomy for treatment of intractable psychiatric disease. *Neurosurgery* 2002; 50: 1043–1052.

Major Depression: Emerging Therapeutics

Srijan Sen, MD, PhD and Gerard Sanacora, MD, PhD

Department of Psychiatry, Yale University School of Medicine, New Haven, CT

ABSTRACT

The first effective antidepressants, monoamine oxidase inhibitors and tricyclic antidepressants, were identified 50 years ago, largely through serendipity. These medications were found to improve mood in a little more than half of depressed patients after a few weeks of chronic use. Almost all antidepressants prescribed today were developed through minor modifications of these original antidepressants and, like monoamine oxidase inhibitors and tricyclic antidepressants, act primarily through monoaminergic mechanisms. Although there have been improvements in side-effect profiles and overdose toxicity, these newer medications have not provided substantial advances in the efficacy and speed of the antidepressant effect for patients. Over the last 2 decades, our understanding of the neurobiology underlying depression has expanded exponentially. Given this expansion, we may be nearing an inflection point in antidepressant drug development, at which useful medicines will be designed through a rational understanding of the biological systems. In this review, we discuss the biological basis and preclinical and clinical evidence for a series of promising classes of antidepressants developed primarily out of a pathophysiologically informed approach. ***Mt Sinai J Med 75:204–225, 2008.*** © *2008 Mount Sinai School of Medicine*

Address Correspondence to:

Gerard Sanacora
Department of Psychiatry
Yale University School of Medicine
New Haven, CT
Email: gerard.sanacora@yale.edu

Key Words: α-amino-3-hydroxy-5-methyl-isoxazole-4-propionic acid, antidepressant, corticotropin-releasing factor, cytokines, glutamate, hypothalamic-pituitary-adrenal, neurotrophic factors, *N*-methyl-D-aspartate, plasticity, melatonin, substance P.

Major depressive disorder (MDD) is characterized by a constellation of symptoms, including sleep and eating disturbances, depressed mood, feelings of worthlessness, increased guilt, and suicidal ideation. MDD affects approximately 16% of Americans at some point in life,[1] and according to the World Health Organization, it will soon be second only to cardiac ischemia in disease-related morbidity and mortality.[2] Yet, despite the continuing burden of the disorder on individuals and society, there has not been a major advance in our pharmacological treatment strategies since the development of the monoaminergic agents [selective serotonin reuptake inhibitors (SSRIs), serotonin-norepinephrine reuptake inhibitors (SNRIs), tricyclic antidepressants (TCAs), and monoamine oxidase inhibitors (MAOIs)], which were developed, in principle, decades ago. These medications act to increase the activity of the monoaminergic neurotransmitters serotonin, norepinephrine, and dopamine. Largely as a result of the mechanisms of these existing medications, most depression research over the past half-century has focused on the neurobiology of the monoaminergic systems. Numerous animal studies, imaging paradigms, genetic studies, and depletion studies have definitively identified vulnerabilities within the serotonin, norepinephrine, and dopamine systems in depression.[3–5] However, it is increasingly clear that the pathophysiological abnormalities associated with MDD go beyond the monoamines alone.

Perhaps the best evidence demonstrating our need to look beyond the monoaminergic systems in our study of MDD is the limited efficacy of our current monoaminergic treatments. The Sequenced Treatment Alternatives To Relieve Depression (STAR*D) trial, a unique large, multicentered effectiveness trial free of industry sponsorship, found that 37% of

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20043

© 2008 Mount Sinai School of Medicine

patients achieved remission with the SSRI citalopram and that only about half of those patients who remained depressed after the trial of citalopram achieved remission with 3 follow-up treatment steps involving medication switching and augmentation.[6] Additional limitations of our current antidepressants include the 2- to 5-week response lag and high rates of relapse.[6] Clearly, there would be great public health benefit from the development of medications with greater efficacy, faster onset of action, and improved tolerability. Moreover, with the increasing advances in pharmacogenomics and personalized medicine, even the development of new classes of medications with efficacy similar to that of existing antidepressants could potentially increase the likelihood that any individual patient would be able to find an effective medication. Thus, there is great interest in the development of new treatments for depression.

Given the complexity of the depression phenotype and the infancy of our understanding of the underlying neurobiology at the time of development, it is not surprising that the monoaminergic antidepressants that currently dominate the market were identified primarily through serendipity rather than rational drug design. Although interest remains in producing new antidepressants through modifications of current medications, the substantial advances in our understanding of the underlying pathophysiology of depression have created great hope that truly novel agents produced through pathophysiologically informed hypotheses are on the horizon. In this review, we explore how advances in our understanding of pathophysiological processes associated with MDD are guiding the development of several novel classes of antidepressant medications.

NEUROTROPHIC FACTORS

One of the major advances in neuroscience over the past 2 decades is the growing understanding of the key roles that neuroplasticity and cell proliferation play in the adult brain.[7] In addition to learning, memory, and neurodegenerative diseases, there is increasing evidence implicating neurotrophic factors in the pathophysiology of depression.[8] Neurotrophins are a class of peptides widely distributed throughout the peripheral and central nervous system that are released synaptically through a calcium-dependent mechanism. Neurotrophins bind to the tropomyosin receptor kinase (TRK) family of receptors (among others), acting to up-regulate a large array of neural signaling pathways.[8] Although many specifics of this network remain to be elucidated, the pathways appear to largely converge to phosphorylate and activate the transcription factor cyclic adenosine monophosphate Ca^{2+} response element binding protein (CREB).

Brain-Derived Neurotrophic Factor (BDNF), Depression, and Stress

Among neurotrophins, the growth factor BDNF has been most extensively studied with respect to the pathophysiology of depression. BDNF is a small dimeric protein found throughout the brain, with particular abundance in the hippocampus and cerebral cortex.[9] Depressed subjects show reduced levels of hippocampal and cortical BDNF in postmortem studies.[10] This is of special interest because hippocampal lesions have effects on anxiety and depressive behaviors,[11] and the hippocampi of depressed patients have been found to be smaller than those of healthy individuals in most but not all studies.[12] Furthermore, hippocampus-specific deletion of BDNF prevents extinction of aversive memories, which may be related to ongoing negative thought processes in depression.[13] In addition, polymorphisms in BDNF have been associated with depression-related traits in some but not all studies,[14–16] and direct infusion of BDNF into the brains of animals has been shown to produce an antidepressant effect in learned helplessness and forced swim models of depression.[17]

As discussed in the hypothalamic-pituitary-adrenal (HPA) section of this review, there is a well-established connection between stress and MDD. A wide range of stressors and the administration of glucocorticoids have been shown to lower transcription of BDNF in the hippocampus.[8] Evidence that chronic social defeat stress induces long-term methylation of histone proteins adjacent to BDNF provides a possible mechanism for the effects of stress on BDNF expression.[18]

Universal Antidepressant Effect

Although several forms of stress have been shown to down-regulate neurotrophic factors, almost all effective depression treatments (SSRIs, SNRIs, TCAs, MAOIs, electroconvulsive therapy, and exercise) induce BDNF expression with chronic administration.[19] Furthermore, these antidepressants induce the phosphorylation of CREB and increase neurogenesis and neuronal turnover in the hippocampus.[20,21] TCAs down-regulate histone deacetylase activity, resulting in increased activity of BDNF, and this indicates that like stress, the effects of antidepressants on BDNF may be mediated, at the genetic level, by histones.[18]

Additional evidence for the role of BDNF in the behavioral effects of antidepressants comes from several lines of neurotrophin-related knockout (KO) mice that have been developed. These strains show a substantially reduced response to antidepressant treatment, and this indicates that increased BDNF activity is necessary to produce the behavioral effects of antidepressants.[22] Together, these lines of evidence indicate that antidepressants universally act to increase neuroplasticity and BDNF activity and that these effects may be necessary for antidepressants to be effective.

Caveats

Although this large body of evidence strongly implicates BDNF in depression, some lines of work caution against the simple interpretation that mood is proportional to BDNF activity. First, KO mice with disruption in the BDNF signaling system do not consistently exhibit depressive behaviors in mouse models of depression.[8] Similarly, antagonism of BDNF activity does not consistently induce depressive behavior.[21] In addition, although BDNF appears to have antidepressant effects in the hippocampus, BDNF infusion into the ventral tegmental area induces depressive behaviors. Along the same lines, intact BDNF activity in the mesolimbic dopamine pathway appears necessary to mediate the depressive affects of social defeat stress.[23] Thus, although it appears clear that BDNF is an important factor in mediating the relationship between neuroplasticity and depression, the role of BDNF may be more complicated than originally thought.

Additional Neurotrophic Factors

Partially because of these complications, there has been increasing interest in other neurotrophic factors beyond BDNF that could mediate neuroplastic changes. For instance, recent work has shown that in addition to BDNF, vascular endothelial growth factor and insulin-like growth factor 1 also show a consistent pattern of up-regulation in response to antidepressants.[24,25] Similarly, dysregulation of fibroblast growth factor proteins has been implicated in depression.[26] It is likely that future work will more clearly elucidate the role of these and other neurotrophic factors in the neuroplastic changes associated with depression.

Clinical Evidence

Although the infusion of BDNF has shown antidepressant properties in animal models, translating this effect into clinically useful medications has proven difficult. In degenerative disease clinical trials, BDNF and other neurotrophic factors showed poor pharmacokinetic properties and low efficacy.[27] Furthermore, development of small-molecule agonists of neurotrophic receptors has been largely unsuccessful to date. A more fruitful strategy may entail targeting elements of the neurotrophic pathways that are more amenable to drug development.[19] For example, neurotrophins act to inhibit the activity of phosphodiesterases, resulting in increased CREB activity. Direct inhibition of phosphodiesterases through the phosphodiesterase 4 (PDE4) inhibitor rolipram has been proved to be effective in depression in clinical trials.[28,29] Although the utility of rolipram is severely limited by side effects, there is continued effort to develop second-generation PDE4 inhibitors with antidepressant efficacy and a tolerable side-effect profile.[30] Another result of neurotrophic activity is inhibition of glycogen synthase 3 (GSK-3). Given that GSK-3 is also directly inhibited by lithium and other treatments effective in mood disorders, it is an attractive target for antidepressant development.[31] GSK-3 inhibitors have shown antidepressant efficacy in animal models of depression, but results from clinical trials have yet to be reported.[31]

PDE4 inhibitors and GSK-3 inhibitors represent just 2 classes of compounds within the neurotrophic system that are under investigation as potential antidepressants. As our understanding of neurotrophic pathways continues to grow, additional potential targets will likely emerge. Although no therapeutics directly targeting the neurotrophic system appear likely to have clinical utility in the immediate future, given the strong preclinical evidence supporting a role in MDD and the large number of potential targets, there is great hope that further exploration of this system will provide clinical benefit in the long term. In addition, the finding that all treatments with antidepressant effects up-regulate neurotrophic activity indicates that this system may prove useful in testing the utility of potential antidepressants derived from other neural systems.

AMINO ACID NEUROTRANSMITTER SYSTEMS

Although most of the attention in psychiatric research has been directed toward the monoaminergic systems, it is important to recognize that the amino acid neurotransmitters, namely glutamate (Glu) and gamma aminobutyric acid (GABA), are the major excitatory and inhibitory transmitters, being responsible for the large majority of all neurotransmitter activity in the brain. Not surprisingly, the amino acid neurotransmitter systems have been found to play critical roles in nearly all forms of learning, emotional processing, and behavior. They are also believed to play critical roles in the regulation of neuroplasticity. However, the potential ability to target these systems for the development of antidepressant drugs has only relatively recently gained true attention. Interest in the amino acid neurotransmitter systems as targets for antidepressant drug development largely results from the convergence of several lines of evidence implicating the systems in the pathophysiology and possible pathogenesis of depression.[32] This pathophysiologically informed approach has led to several hypotheses in which modulation of amino acid neurotransmission may have antidepressant-like properties. We briefly review the preclinical work that supports these hypotheses and outline some of the early phase clinical trials that are now beginning to appear in the literature.

Glutamatergic Drugs

Several lines of evidence suggest that abnormalities in the glutamatergic neurotransmitter system contribute to the neurobiology of MDD (see Kugaya and Sanacora,[32] Kugaya *et al.*,[33] and Pittaluga *et al.*[34] for comprehensive reviews). In nonhuman studies, there is evidence that stress is associated with increased levels of glutamatergic activation in several brain regions.[35] This increased glutamatergic drive is believed to contribute to various forms of stress-related neurotoxicity.[36] Other studies have recently demonstrated that repeated stress significantly decreases the number and volume of glial cells[37] and reduces the rate of glial cell proliferation.[38] The importance of these findings derives from the fact that glial function is critical in the regulation of both glutamatergic and GABAergic neurotransmitter systems, and there is a relatively large series of postmortem studies now demonstrating decreased glial cell number and density in various brain regions in depressed individuals (see Kugaya and Sanacora[32] and Rajkowska and Miguel-Hidalgo[39] for reviews and related details).

Other studies have provided evidence of altered Glu function in depressed individuals. Several postmortem studies have demonstrated abnormalities in the expression and binding of glutamate receptors (GluRs) and the proteins in the postsynaptic membrane that are responsible for anchoring, trafficking, and providing a means of signal transduction for the receptors.[40–44] In vivo magnetic resonance spectroscopy studies have demonstrated abnormal Glu content in various brain regions of individuals diagnosed with MDD.[45–50] In sum, findings from both basic science and clinical laboratories are now providing strong evidence that abnormalities within the Glu neurotransmitter system are associated with MDD. These findings have served to fuel the interest in the development of pharmaceutical agents targeting the glutamatergic system for the treatment of MDD. Fortuitously, as a result of the complex regulatory processes controlling Glu signaling, the system provides an abundance of potential targets for pharmacological modulation. Next, we briefly cover several of the potential targets that are being examined in relationship to the treatment of MDD.

Targeting GluRs

There are 3 excitatory ionotropic GluRs, which are identified by their pharmacology as *N*-methyl-D-aspartate (NMDA), α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA), and kainate (KA), and there are several metabotropic glutamate receptors (mGluRs), which are divided into 3 groups primarily on the basis of their coupling to different signal transduction systems and ligand sensitivities. Emerging evidence suggests that several of these receptor families could serve as potential targets for antidepressant drug development (Figure 1).

NMDA Receptors

The NMDA receptor has gained considerable attention throughout the field of neuroscience because of its roles in learning, neuroplasticity, and neurotoxicity. The receptor has several unique properties, such as the need for a coagonist (glycine), a voltage-sensitive block of the ion channel by magnesium, the presence of multiple allosteric binding sites, and interactions with various scaffolding proteins that mediate the transport and localization of the receptor as well as interactions with multiple signaling, which make it well suited to playing a critical role in regulating neuroplastic changes. These same properties also make it a favorite target for the development of pharmaceutical interventions. NMDA receptors are believed to exist primarily as tetrameric complexes composed of 2 obligatory NR1 subunits and 2 NR2 subunits. At least 8 splice variants of the NR1 subunit exist, along with 4 NR2 genes (NR2A, NR2B, NR2C, and NR2D) and 2 NR3 subunits (NR3A and NR3B). The receptors are mobile and move laterally between the synaptic and extrasynaptic pools, and recent evidence suggests that the location of the receptor with respect to the synaptic space may have large effects on NMDA receptor signal transduction. Specifically, it appears that synaptic NMDA receptors activate mitogen-activated protein kinase and the transcription factor CREB, induce

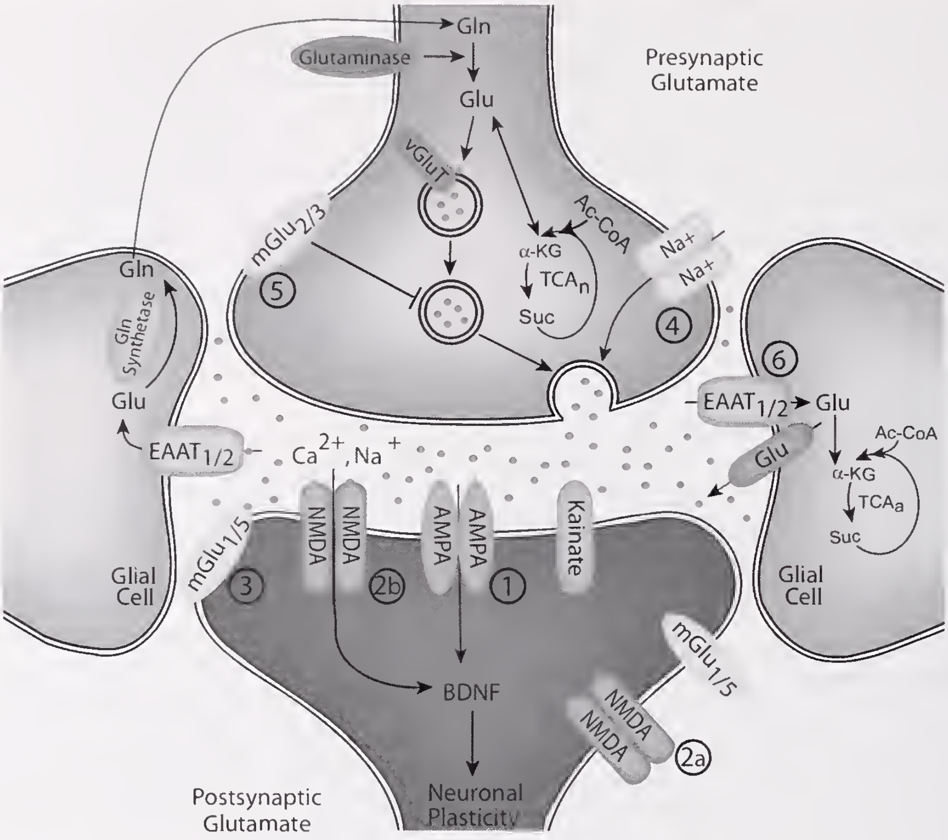


Fig 1. Potential glutamatergic targets for drug development: (1) AMPA receptor potentiators, (2) NMDA receptor antagonists [the potential to target the various NR2 subunits may provide a differential ability to modulate (2a) extrasynaptic and (2b) synaptic NMDA receptors], (3) mGluR1/5 antagonists, (4) inhibition of stimulated Glu release via voltage-dependent Na+ channels, (5) mGluR2/3 receptor agonists (they have been shown to have anxiolytic properties, whereas mGluR2/3 receptor antagonists have demonstrated antidepressant activity), and (6) facilitation of Glu clearance by EAATs. **Abbreviations:** ±-KG, Alpha-ketoglutarate; Ac-CoA, acetyl coenzyme A; AMPA, α-amino-3-hydroxy-5-methyl-isoxazole-4-propionic acid; BDNF, brain-derived neurotrophic factor; EAAT, excitatory amino acid transporter; Gln, glutamine; Glu, glutamate; mGlu, metabotropic glutamate; mGluR, metabotropic glutamate receptor; NMDA, *N*-methyl-D-aspartate; Suc, succinic acid; TCA, tricyclic antidepressant; vGluT, vesicular glutamate transporter. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

expression of the gene that encodes BDNF, and promote neuronal survival, whereas extrasynaptic NMDA receptors propagate opposing signals that promote cell death.[51,52] An emerging concept posits that activity-dependent, bidirectional regulation of NMDA receptor trafficking provides a mechanism for the regulation of synaptic efficacy and remodeling, which, if dysregulated, can contribute to several neuropsychiatric disorders.[53]

Several early reports by Skolnick,[54] Paul *et al.*,[55] and others suggested that the receptor may be a target of antidepressant action by showing that a drug's ability to alter the NMDA receptor's binding affinities was highly correlated to its antidepressant activity in the clinic. Interest in the NMDA receptor grew rapidly after the serendipitous finding that a single subanesthetic dose (0.5 mg/kg over 40 minutes) of ketamine, a noncompetitive NMDA receptor antagonist, produced a rapid and sustained response in 7 individuals with previously treatment-resistant MDD.[56] Zarate and others at the National Institute of Mental Health[57] recently replicated the finding in a larger study of 18 subjects, which found more than 70% of the subjects to have an antidepressant response to ketamine within 24 hours of the infusion. The study also demonstrated that nearly 30% of the subjects benefited to the level equated with remission by the end of the first day. Together, the studies consistently demonstrated ketamine's ability to induce a rapid improvement in depression severity. Although a large degree of variability in the duration of the response between individuals was noted, there was a strong tendency for rating scores to begin to return to the pretreatment levels by the fifth day post-infusion for most subjects.

Although the results of these studies have generated great interest in the potential development of NMDA antagonists as antidepressant agents, ketamine produces several acute effects, such as altered levels of consciousness, perceptual and cognitive disturbance, and hypertension, that significantly limit its potential usefulness in the clinic setting. The intravenous route of administration used in the previous clinical trials may further limit its widespread use; however, this may become less of an issue as intranasal preparations of ketamine are now available.[58] Perhaps the most important consideration limiting the clinical use of ketamine or any other NMDA antagonist in the treatment of MDD is that it has not yet been shown that repeated doses of drugs will have the same antidepressant-like response. The fact that the drug produces a rapid improvement in mood after a single administration does not necessarily mean that chronic administration will sustain the response. In fact, it is possible that chronic administration of NMDA antagonists could have detrimental effects on neuronal survival,[59] although the evidence of this occurring in primates, especially outside of the perinatal period, remains controversial.[60] Although these concerns should not preclude further exploration of the clinical applications of ketamine in the treatment of MDD, it is clear that they will need to be addressed before large-scale clinical trials using chronic dosing regimens can proceed with ketamine.

One way of attempting to address these concerns has been to identify agents acting as NMDA subunit–selective antagonists. For example, drugs such as ifenprodil have been shown to be highly selective for the NR2B subunit–containing NMDA receptors. These compounds appear to have a lower liability for these unwanted effects and thus may provide greater clinical utility. CP-101–606, another NR2B-specific NMDA antagonist, was recently reported in abstract form to produce a statistically significant improvement in depressive symptoms in a double-blind, randomized, placebo-controlled parallel group study with subjects with recurrent MDD.[61] The average reduction in symptom severity was approximately 2 times greater in actively treated patients versus placebo-treated patients, whereas dissociative symptoms produced by the active infusion were generally modest and resolved within 8 hours of administration. Additional studies with these selective compounds will be required to determine the optimal subunits to target and the true clinical utility of this approach.

AMPA and KA Drugs

AMPA receptors mediate the fast, rapidly desensitizing excitation at the majority of synapses and are responsible for the initial reaction to Glu in the synapse. Similar to the NMDA receptor, the AMPA receptors are composed of a homomeric or heteromeric complex of 4 subunits (GluR1, GluR2, GluR3, and GluR4). Differences in individual subunit expression provide a large degree of functional diversity to this class of receptors. AMPA receptors are usually coexpressed with NMDA receptors at mature synapses, where together they contribute to the processes of synaptic plasticity that are involved in memory, learning, excitotoxicity, and neuroprotection.[62,63] Modification of the levels of synaptic AMPA receptors, particularly by receptor subunit trafficking, insertion, and internalization, is now recognized to be a critically important mechanism for regulating various forms of synaptic plasticity.[62] KA receptors have diverse presynaptic and postsynaptic roles and have some of the same properties as AMPA receptors.[64] The KA receptors are coded by 2 gene families, which code for the low-affinity GluR5-7 subunits and the high-affinity KA1 and KA2 subunits. Like AMPA receptors, KA receptors are associated with voltage-dependent channels that mediate fast excitatory neurotransmission. However, they appear to have a distinctive distribution and unique functions.[65]

Both the AMPA and KA receptors have gained increasing notoriety with respect to the area of mood disorder research because of preclinical studies suggesting that the receptors are altered in animal models of mood disorders and in depressed individuals.[66] Additional evidence shows that AMPA receptor function is modified by many classes of existing mood stabilizers and antidepressant agents, thus suggesting a potential common mechanism of action.[67] In fact, a recent study suggests that the functional activation of the AMPA receptor may be a necessary component of the antidepressant effects of ketamine described previously.[68] Interest in the AMPA and KA receptors with respect to MDD has been further spurred by 2 recent reports associated with the STAR*D clinical trial that were designed to explore the association between genetic variation and outcome following citalopram treatment. The first study suggested that allelic variations in the glutamate receptor ionotropic kainate 4 gene, which codes for the KA1 subunit of the KA receptor, is associated with nonresponse to treatment and a subtype of anxious depression.[69] The second study demonstrated that genes coding for the GluR3 AMPA receptor subunit and the GluR6 KA receptor subunit appear to be associated with suicidal ideation after treatment with citalopram.[70] In a potentially important step toward the advancement of personalized medicine in psychiatry, a biotech company (NeuroMark) has already marketed a version of this genetic test for clinical use, as reported in a recent media report (http://abcnews.go.com/technology/story?id=3673121@page=1).

With respect to drug development, recent studies have demonstrated that several compounds that augment signaling through AMPA receptors (AMPA receptor potentiators) and both AMPA and KA receptors (AMPAkines) exhibit antidepressant-like behavioral effects in animal models. They also produce neuronal effects similar to those produced by currently available antidepressants, including neurotrophin induction and increases in hippocampal progenitor cell proliferation. This suggests that several classes of compounds that target AMPA receptors, including pyrrolidones (piracetam and aniracetam), benzothiazides (cyclothiazide), benzylpiperidines (CX-516 and CX-546), and biarylpropylsulfonamides (LY392098, LY404187, LY450108, LY451395, and LY503430), may prove useful in the clinical management of depression (see Bleakman *et al.*,[66] Alt *et al.*,[71] and O'Neill and Witkin[72] for more complete reviews). However, no clinical data on any of these compounds related to depression or mood disorders are yet available.

mGluR Drugs

mGluRs are G protein–coupled receptors that are divided into 3 groups on the basis of sequence similarity, effector coupling, and ligand sensitivity: members of group I class (mGluR1 a-d and mGluR5 a-b) primarily act through phospholipase Cb and the activation of the inositol trisphosphate and diacylglycerol second messenger systems.[73] They are located mostly on postsynaptic neurons at the fringe of the postsynaptic density, where they are believed to play a major role in regulating postsynaptic excitability. However, they are also present on glial cells, where they appear to regulate the Ca^{2+}-dependent release of excitatory transmitters from these glia.[74] Groups II (mGluR2 and mGluR3) and III (mGluR4, mGluR6, mGluR7, and mGluR8) are negatively coupled to adenylyl cyclase and are largely present on the presynaptic membrane, where they are likely to respond to Glu spillover by regulating transmitter release.

Preclinical studies have provided evidence that group I metabotropic receptor modulating agents, especially mGluR1/5 antagonists such as 2-methyl-6-(phenylethynyl) pyridine and 3-[(2-methyl-1,3-thiazol-4-ylethynyl]pyridine, have strong anxiolytic activity in preclinical animal models. More recent studies have suggested that they may also possess antidepressant-like properties (see Palucha and Pilc[75] and Witkin *et al.*[76] for recent reviews on this topic). However, no clinical studies demonstrating the effectiveness of this class of drugs have been reported to date.

Group II metabotropic receptors (mGluR2 and mGluR3) are largely present on the presynaptic membrane and on glial cells, where they are believed to modulate Glu neurotransmission by sensing Glu spillover and regulating transmitter release. Recent work has suggested that the group II metabotropic receptors also serve an important role in the mechanisms of neuroprotection and neurodegeneration.[77] The function of the mGlu2/3 receptors in modulating Glu neurotransmission makes them extremely interesting targets for antidepressant drug development, and rapidly accumulating evidence suggests that drugs targeting the mGluR2/3 receptors possess both antidepressant and anxiolytic properties (see Palucha and Pilc[75] and Witkin *et al.*[76] for reviews). However, the picture is complicated by the fact that mGluR2/3 receptor antagonists (eg, LY341495 and MSG0039) appear to have antidepressant-like properties,[78] whereas mGluR2/3 receptor agonists (ie, LY404039 and LY354740) have a profile consistent with anxiolytic drugs. Similarly to the case with the NMDA antagonists, many of the antidepressant-like effects of the mGluR2/3 receptor antagonists can be prevented by coadministration of an AMPA receptor antagonist,[79] and this suggests that activation of synaptic AMPA receptors may be a common pathway through which the antidepressant effects are transduced. Early phase clinical trials have recently supported the efficacy of mGluR2/3 receptor agonists in the treatment of anxiety disorders but raise some concern over the safety of this class of drugs.[80] A recent randomized, double-blind, placebo-controlled study also suggested that mGlu2/3 receptor agonists (LY2140023) have antipsychotic properties and may provide a new alternative for the treatment of schizophrenia.[81] However, no clinical studies have yet been conducted to explore the antidepressant efficacy of mGluR2/3 receptor drugs.

The evidence of Glu abnormalities in mood disorders has also motivated the investigation of other medications that modulate Glu neurotransmission by less clearly defined mechanisms. Because of its unique Glu-modulating properties, the medication riluzole (Rilutek), a drug currently Food and Drug Administration–approved and marketed for the treatment of amyotrophic lateral sclerosis,[82] has gained considerable attention. Riluzole (2-amino-6-trifluoromethoxy benzothiazole) appears to modulate Glu neurotransmission by several pharmacological mechanisms. The drug was originally developed as an anticonvulsant, and early studies showed it to oppose several actions of Glu in vitro and in vivo.[83,84] Various studies have demonstrated that the drug may act through functional antagonism at a subset of GluRs[83,85] and through interactions with a number of voltage-activated sodium,[86–89] calcium,[90,91] and potassium channels[88,92–95] that result in a reduction in Glu release.[96–99] Recently, riluzole has also been found to enhance glial cell reuptake of Glu[100–102] and increase the surface expression of the GluR1 and GluR2 subunits of AMPA at the synapse[67]; these 2 mechanisms may prove to be of substantial importance for its clinical actions. Similar to many other antidepressants, riluzole demonstrates the ability to increase hippocampal BDNF and cell proliferation.[103]

Following an initial case report of a patient with comorbid obsessive compulsive disorder and MDD who appeared to experience a clinical benefit in both mood and obsessive-compulsive disorder symptoms after starting riluzole,[104] there have been several additional reports of open-labeled clinical trials suggesting that riluzole possesses potent antidepressant and anxiolytic properties. An initial open-label clinical trial examining the effects of riluzole in 19 treatment-resistant depressed patients found a significant improvement in mood ratings beginning in the third week of treatment.[105] The same investigators also conducted an open-label trial of riluzole augmentation of lithium treatment in bipolar subjects with a treatment-resistant depressive episode.[106] This study found significant improvement in depression ratings with no reports of hypomania or mania. A more recent study demonstrated that riluzole augmentation of traditional antidepressant medications produced a significant improvement in depressive symptoms in 10 subjects with severe treatment-resistant depression.[107] Moreover, subjects who achieved a remission in the initial 6-week phase of this study maintained their remission for the 3-month duration of the study. Another recent open-label study suggests that riluzole also has efficacy in the treatment of generalized anxiety disorder.[108] In this 8-week, open-label, fixed-dose study, 18 medically healthy patients with generalized anxiety disorder (according to the *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed.) received treatment with riluzole following a 2-week drug-free period; 12 of the 15 patients who completed the trial responded positively to riluzole, and at 8 weeks, 8 of the 15 patients had Hamilton Rating Scale for Anxiety scores indicating remission of their anxiety.

Together, these studies suggest that riluzole, a unique Glu-modulating agent, has significant antidepressant and anxiolytic effects in subjects with MDD resistant to standard monoamine-focused therapies and is well tolerated. However, the total number of patients who have been described in the 4 trials remains small, and no double-blind studies have been described to date.

GABA System

The GABAergic system has become increasingly implicated in the neurobiology of mood disorders. Supporting evidence of GABAergic involvement in the pathophysiology and treatment of mood disorders is supported by several lines of evidence, including (1) animal studies showing stress-related changes in GABAergic function, (2) the ability of GABA agonists and antagonists to modulate behavioral models of depression in rodents, (3) GABAergic effects of standard antidepressant medications, and (4) the demonstration of GABAergic abnormalities and genetic associations in depressed patients (see Sanacora and Saricicek[109] for a complete review). Based on the evidence of GABAergic abnormalities in depressed subjects and the potent effects of GABAergic drugs in alleviating anxiety, several studies have attempted to explore the potential effectiveness of GABAergic drugs in the treatment of mood disorders. In general, the results have been mixed, with most familiar benzodiazepine compounds not proving effective as monotherapeutic agents in the treatment of MDD.[110] Identification of subunit-specific behavioral and physiological effects has focused the field on the development of subtype-selective compounds. Studies demonstrating the sedating and anticonvulsant properties of $GABA^A$ receptors containing the $\alpha1$ subunit have successfully led to the development of several more selective agents (eszopiclone, zaleplon, zolpidem, and indiplon) for use as hypnotic agents. Other work suggesting that the $\alpha2$ and $\alpha3$ subunit–containing receptors are associated with the regulation of anxiety[111–113] has led to the development of selective 2,3-agents[112] for the treatment of irritable bowel syndrome and anxiety. Although there is preliminary evidence suggesting that eszopiclone may augment the antidepressant effects of standard antidepressant agents,[114] there is no clear evidence suggesting antidepressant efficacy of these drugs as monotherapeutic agents. Although there is now strong evidence demonstrating that GABAergic dysfunction is associated with mood disorders, there has not been great success in targeting the system for the development of antidepressant drugs (see Sanacora and Saricicek[109] for a recent comprehensive review).

HPA AXIS

The hypothalamus, pituitary, and adrenal glands together compose the HPA axis, a highly regulated hormonal system that plays an integral role in mediating stress response. Dysregulation of the HPA axis has been one of the most consistent and heavily investigated findings in depression research.[115] Depressed patients show excessive cortisol secretion at baseline and inadequate inhibitory feedback regulation of the HPA axis. Furthermore, successful treatment of depression, regardless of modality, corrects the HPA dysregulation.[116,117] Given its relationship to stress and depression, the HPA axis is a natural target for antidepressant medications (Figure 2).

Corticotropin-Releasing Factor (CRF) Antagonists

Within the HPA axis, perhaps the most promising progress to date involves CRF. CRF (previously called corticotropin-releasing hormone) is a 41 amino acid protein that is released from neurons in the paraventricular nucleus of the hypothalamus and stimulates the release of adrenocorticotropic hormone (ACTH) from the anterior pituitary. ACTH, in turn, promotes the release of cortisol from the adrenal cortices.

A growing body of evidence indicates that increased CRF activity is the primary driver of the HPA system dysregulation found in depression.[118] Transgenic mice overexpressing CRF show phenotypic behaviors consistent with depression.[119] Similarly, the injection of exogenous CRF into the lateral ventricle of mice results in decreased appetite, increased anxiety, and other depressive behaviors.[120,121] Interestingly, these depressogenic effects of CRF appear to be potentiated in the context of external stress such as an unfamiliar environment or social defeat.[119] CRF levels are elevated in the cerebrospinal fluid (CSF) of depressed subjects; this abnormality is corrected by a number of antidepressant therapies.[122] In addition, people who suffer early life stress show elevated central CRF levels as adults.[123,124] Furthermore, the link between early life stress and adult depression appears to be mediated in part by variations in the CRF1 receptor gene.[229] The release of ACTH in response to CRF administration is blunted in depressed subjects.[125] Suicide victims show an increased level of frontal CRF in comparison with control subjects.[118] Transgenic mice overexpressing CRF show phenotypic behaviors consistent with depression.[119] Similarly, the injection of exogenous CRF into the lateral ventricle of mice has a depressogenic effect.[126] In addition to mediating the endocrine stress response, CRF appears to be involved in autonomic and immunologic responses as well.[127,128]

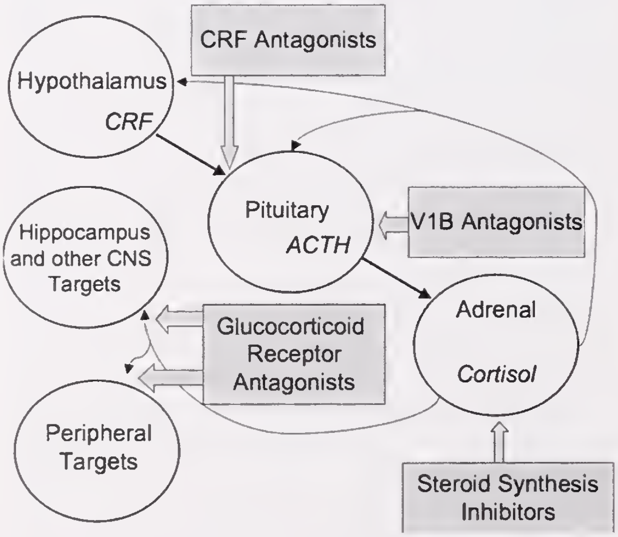

CRF Antagonists
Hypothalamus
CRF
Pituitary
ACTH
V1B Antagonists
Hippocampus and other CNS Targets
Adrenal
Cortisol
Glucocorticoid Receptor Antagonists
Peripheral Targets
Steroid Synthesis Inhibitors

Fig 2. Novel antidepressant targets within the hypothalamic-pituitary-adrenal system. **Abbreviations:** ACTH, adrenocorticotropic hormone; CNS, central nervous system; CRF, corticotropin-releasing factor; V1B, vasopressin 1b. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

With this growing base of evidence, the antagonism of CRF has emerged as a novel target for antidepressant drug development. CRF activates 2 receptor subtypes, CRF1 and CRF2. Evidence from rodent models indicates that the CRF1 receptor is the primary mediator of the CRF-mediated stress response. CRF1 receptors are aggregated in the pituitary (the site of ACTH release) as well as the amygdala, hippocampus, cerebellum, and prefrontal cortex. CRF1 KO mice show decreased anxious and depressive behaviors.[129,130] Furthermore, antagonism of CRF1 reduces the anxiogenic effects of stress and exogenous CRF administration.[131] Interestingly, mice with decreased CRF1 function in the amygdala, hippocampus, and cortex but intact pituitary CRF1 function show reduced anxiety, and this indicates that at least some of the anxiogenic effects of CRF1 are mediated independently of the HPA axis.[230]

In contrast to the clear anxiogenic role of CRF1, the role of the CRF2 receptor in anxiety and stress response is more ambiguous. In comparison with CRF1, CRF2 receptors have an approximately 15-fold lower affinity for CRF. CRF2 receptors are found mostly in the lateral septum and hippocampus, with lower density within the HPA axis.[132] Despite the general anxiogenic effects of increased CRF activity, CRF2 KO mice display increased anxiety, and exogenous administration of a CRF2 agonist decreases anxiety.[133] In contrast, the central infusion of a CRF2 antagonist and antisense inhibition of CRF2 also produce anxiolytic effects. These apparently contradictory findings may be a result of differential effects of the receptor in different brain regions, but much work remains to clarify the role of CRF2 receptors.[134] Given the ambiguity surrounding CRF2 receptors, antagonism of CRF1 has emerged as the primary target within the CRF system for antidepressant medications.

Preclinical Models

Several CRF1 antagonists have been examined in learned helplessness, forced swim, and tail suspension animal models of depression.[135] Across all 3 paradigms, CRF1 antagonists showed antidepressant effects in some but not all trials.[136–138] Although there are exceptions, 2 general themes have emerged from these trials. First, CRF1 antagonists appear to have antidepressant effects after chronic but not acute dosing. This is a similar property found in most established antidepressants, both in rodent models and in patients. The second salient variable explaining the discrepancy between CRF1 antagonist trials may be the baseline level of anxiety of tested individuals. For example, in comparison with a line of low–baseline anxiety rats (Flanders resistant line), high–baseline anxiety rats (Flanders sensitive line) show a markedly stronger response to CRF1 antagonists.[139] This constellation of findings is consistent with the more general finding that the effects of CRF activity are potentiated in the context of stress and anxiety.

Clinical Trials

Backed by a strong pathophysiological model and promising preclinical data, CRF1 antagonists have moved to human trials with great expectations.[140] Zobel and colleagues[141] investigated the CRF1 antagonist R121919 in a placebo-controlled trial of 20 subjects. The investigators found a significant reduction in anxiety and depressive scores in both patient and clinician ratings. A subsequent investigation raised concerns, however, that the medication was associated with a transient elevation of liver function enzymes, resulting in the discontinuation of its development. A number of pharmaceutical companies are currently investigating alternate CRF1 antagonists. Although some promising results are beginning to emerge from these studies,[142] there has also been some disappointment with the speed of progress in this area. In summary, there is strong theoretical and neurobiological evidence supporting CRF1 antagonists in the treatment of depression. The emergence of support from preclinical and clinical research, though promising, is still in its preliminary stages.

Vasopressin 1b Antagonists

In addition to CRF receptors, investigators are exploring the utility of agents that target other components of the HPA axis as potential antidepressants. Vasopressin is a small neuropeptide released in the hypothalamus that appears to play a role in potentiating the actions of CRF.[143] In animal models, the central infusion of vasopressin produces an anxiogenic effect, whereas the infusion of vasopressin receptor antisense has an anxiolytic effect.[144,145] Brattleboro rats, lacking endogenous vasopressin, show minimal differences from wild-type mice at baseline but show an abnormal response to ACTH in response to stress.[146] With respect to antidepressant-like effects, vasopressin 1b receptor KO mice show some alterations in prepulse inhibition but no significant differences in mouse models of depression (forced swim test) and anxiety (elevated plus maze).[147] In contrast, the administration of a vasopressin 1b receptor antagonist produces antidepressant-like responses in forced swim and chronic mild stress models of depression.[148,149] Furthermore, like most established antidepressants, vasopressin 1b antagonists result in increased neurogenesis.[149] In humans, genetic variation in the vasopressin 1b receptor gene is associated with depression,[150] and postmortem analysis has demonstrated increased vasopressin-expressing neurons in depressed subjects.[151] In addition, depressed subjects show elevated CSF vasopressin levels that can be corrected with fluoxetine treatment.[151] Vasopressin 1b antagonists have now moved to phase II clinical trials, but no evidence of their efficacy is currently available.

Steroid Synthesis Inhibitors

Metyrapone and ketaconazole are inhibitors of cortisol synthesis. In preclinical models, metyrapone, like most established antidepressants, has been shown to induce the transcription of hippocampal BDNF.[152] In clinical trials, although investigators have found mixed results with ketaconazole,[153] both a placebo-controlled, single-blind crossover study and a recent placebo-controlled, double-blind trial found encouraging results for metyrapone as an adjunctive agent for major depression.[154,155] Studies confirming these initial encouraging clinical trials and studies further exploring the efficacy of metyrapone as monotherapy have yet to be reported.

Glucocorticoid Receptor Antagonists

There is compelling evidence that depressed subjects have a mismatch between the 2 primary cortisol receptors, with underactivity of the inhibitory mineralocorticoid receptor and overactivity of the excitatory glucocorticoid receptor.[156] On the basis of the concept that excessive stimulation of the glucocorticoid receptor is associated with more severe forms of MDD, mifepristone (RU486), a strong antagonist of the type II glucocorticoid receptor, has been explored as an adjunctive treatment for psychotic depression.[157] Although the drug has shown efficacy in improving symptoms in some clinical trials and case series,[158,159] pivotal phase III trials for MDD with psychotic features failed to meet their primary depression endpoints[160] but did show efficacy in reducing the Brief Psychiatric Rating Scale scores.[159] In sum, the potential utility of mifepristone in major depression remains unclear at present, but it may have specific use for psychotic symptoms.

POTENTIAL ROLE OF PROINFLAMMATORY CYTOKINES

Driven in part by the increasing awareness of the comorbidity of mood disorders and illnesses known to involve immune system dysfunction, increasing attention has turned to the role of proinflammatory cytokines in the pathophysiology of depression. Cytokines are small intracellular signaling proteins produced by lymphocytes, leukocytes, and glial cells that are necessary in mediating host defense and regulating immune responses. It is now well accepted that there is abundant cross-talk between the immune system and the central nervous system. Circulating proinflammatory cytokines such as interleukin-6 (IL-6) and tumor necrosis factor alpha (TNFα) are believed to mediate much of this communication. Specifically related to the pathophysiology of mood disorders, these factors are known to stimulate corticotropin-releasing hormone, which activates the hypothalamic pituitary axis and thus causes hypercortisolemia,[161] and to act synergistically with Glu to cause neurodegeneration.[162–165]

Acute administration of TNFα or IL-1 either peripherally or centrally induces anhedonia, reduced locomotor and sexual activity, and decreased social interaction in rodents.[166] These symptoms are collectively known as *sickness behavior*, but in humans they are also characteristics of major depression. Conversely, elevated levels of several cytokines have been reported in patients diagnosed with MDD. For example, serum TNFα levels are significantly increased in depressed patients.[167–169] A similar result was obtained in a study examining the association between depressed mood in older adults and plasma levels of IL-6, TNFα , and C reactive protein, an acute phase reactant.[170] Interestingly, these elevated cytokine levels may be normalized by the administration of certain antidepressants, but only in those patients whose depression responded to the treatment.[167,171,172] In an attempt to further explore the relationship between TNFα and depression, Simen *et al.*[173] studied the effect of the TNFα receptor deletions on behavior in mouse models. Both TNFR1 and TNFR2 deletions led to antidepressant-like responses in comparison with wild-type littermates. On the basis of these findings, it is reasonable to hypothesize that cytokines, especially TNFα and IL-6, may contribute to the pathophysiology of MDD.

TNFα antibodies and soluble receptor pharmaceuticals have proven efficacious in the treatment of several autoimmune disorders such as psoriasis, Crohn's disease, ankylosing spondylitis, and rheumatoid arthritis. A recently published double-blind study of 618 subjects treated with etanercept, a soluble TNFα receptor, for moderate to severe psoriasis suggests that the drug may also possess antidepressant activity. The study demonstrated that a significantly greater number of subjects treated with etanercept had at least a 50% improvement on both the Beck Depression Inventory and the Hamilton Depression Rating Scale. The level of improvement appeared to be largely independent of the drug's effects on joint pain and skin involvement.[174] The results of these studies have continued to stir interest in this avenue of research, and clinical trials with a TNFα antibody, infliximab, are currently underway (http://www.clinicaltrials.gov/ct/show/NCT00463580?order=1). Although these TNFα antagonists may hold promise in the treatment of MDD, the routes of delivery and adverse effect profiles may limit their use to more severe, treatment-resistant cases.

SUBSTANCE P (SP) AND NEUROKININ RECEPTORS

SP is an 11 amino acid neuropeptide that is part of the neurokinin family of proteins. Although SP can bind to all 3 G protein–coupled neurokinin receptors, it shows the strongest affinity for neurokinin-1 receptor (NK1R). Both SP and NK1R are found in relatively high concentrations in the amygdala, hippocampus, and hypothalamus, brain regions strongly implicated in depression and stress response.[175–177] In light of this anatomic distribution and the well-established role of SP in pain response, investigators have explored a potential role for the antagonism of SP as a target in depression.

Preclinical Studies

A diverse body of evidence from rodent studies implicates SP in anxiety and depressive behaviors. The injection of SP into the intracerebral ventricle and amygdala of rodents has anxiogenic effects in the elevated plus maze and other models of anxiety.[178,179] Consistent with the idea that SP is associated with depressive symptoms, some but not all studies indicate that chronic treatment with antidepressants results in a decrease in SP concentration,[180,181] and both SP (TAC1-Tachykinin 1) and NK1R KO mice show decreased depressive behaviors in the forced swim and tail suspension models of depression.[182,183]

A series of experiments have also linked SP to neural systems previously implicated in depression. As discussed previously, a strong link between neurotrophic factors, neurogenesis, and depression has emerged. NK1R KO mice show increased hippocampal neurogenesis in comparison with wild-type mice. In addition, these KO mice do not show the typical increase in neurogenesis found with SSRI treatment, and this indicates that SP NK1R activation may be necessary for antidepressant-mediated neurogenesis. In a similar vein, although chronic social defeat stress generally suppresses hippocampal neurogenesis, treatment with NK1R antagonism blocks this stress effect in tree shrews.[184]

In addition to the neurotrophic pathway, SP also appears closely tied to monoaminergic systems previously implicated in depression. Antagonism of NK1R results in increased concentration and activity of norepinephrine in the frontal cortex and hippocampus.[185] Disruption of NK1R also acts to increase firing of serotonergic neurons and desensitization of serotonin neuron autoreceptors, although curiously this does not appear to increase serotonergic neurotransmission.[186–188]

Clinical Studies

Although the bulk of evidence from animal studies implicates increasing SP activity in the pathophysiology of depression, the results of human studies to date are somewhat mixed. Some investigators have found increased SP activity in the CSF of depressed subjects, whereas others have not.[189–191] Similarly, a promising study suggested that a decrease in serum SP levels is predictive of antidepressant response, but again, these findings have not been consistently replicated.[192,193]

Despite these equivocal studies, an early clinical trial of the NK1R antagonist was encouraging and resulted in great interest. In the randomized, double-blind study, Kramer *et al.*[187] found the NK1R antagonist aprepitant to have an efficacy similar to that of paroxetine and significantly greater efficacy than placebo in the treatment of MDD, with no significant side effects. After equivocal results in relatively small follow-up studies, Keller and colleagues[194] conducted a series of 5 clinical trials using about 2500 patients and found no evidence that aprepitant was an effective antidepressant. In light of this result, pharmaceutical company development of NK1R antagonists has slowed considerably. Interestingly, aprepitant has reached the market, but as an antiemetic rather than as an antidepressant. In summary, although there is substantial neuroanatomic, physiological, and behavioral preclinical evidence supporting an important role for SP in the pathophysiology of depression, the preponderance of human studies do not support the use of NK1R receptor antagonists in depression. Although investigators continue to work on NK1R antagonists, some have shifted their focus to an alternate target, the neurokinin system.

MONOAMINES

Although the focus of this review is emerging antidepressants that work through novel mechanisms, it is important to note that there is extensive work taking place both to identify medications approved for other disorders that are effective in depression and to modify existing antidepressants to produce medications with improved efficacy and limited side effects. In terms of existing medications, the most intense interest has revolved around the use of atypical antipsychotics in depression. In addition to their actions on the dopaminergic system, atypical antipsychotics may also act to increase serotonergic tone through actions at serotonin 2A and serotonin 1A receptors. Numerous case reports and, more recently, a series of clinical trials have shown efficacy for aripiprazole, ziprasidone, olanzapine, and risperdal as augmentation therapies for treatment-resistant depression.[195–197] However, the efficacy of these agents as monotherapies is unclear at present. Furthermore, the serious metabolic and weight-gain side effects associated with many atypical antipsychotics could limit their widespread use.

The utility of atypical antipsychotics and the efficacy of the dopamine reuptake inhibitor bupropion, both as monotherapy and as an augmenting agent for serotonergic/norepinephric agents, have raised interest in the development of triple reuptake inhibitors, medications that act to inhibit the reuptake of dopamine as well as serotonin and norepinephrine.[198] In some sense, this development is a conceptual return to some of the first antidepressants, such as TCAs and MAOIs, which may derive their effects through a broad spectrum of actions. In preclinical models, 2 triple reuptake inhibitor compounds have been shown to produce antidepressant effects in forced swim and tail suspension mouse models of depression.[199,200] Multiple triple reuptake compounds have been tested in phase I clinical trials. Side effects of nausea and diarrhea were reported, but otherwise the medications showed good tolerability. In the only phase II trial of which we are aware, a triple reuptake inhibitor developed by DOV Pharmaceuticals showed significantly greater efficacy than placebo but efficacy similar to that of citalopram. At present, other triple reuptake inhibitor compounds are in phase II trials.

OTHER MECHANISMS OF INTEREST

Melatonin

Melatonin, a tryptophan-derived hormone released primarily from the pineal gland, is a key regulator of sleep/wake cycles.[201] The synthesis and release of melatonin are synchronized to a 24-hour cycle through light exposure.[202] Given the central role of sleep dysregulation and circadian rhythms in depression, it is not surprising that melatonin has been heavily investigated in depression research. Both the concentration of melatonin and the timing of melatonin release are perturbed in depression.[203,204] Furthermore, chronic treatment with various classes of existing antidepressants results in increased melatonin levels.[202,205]

In general, the administration of melatonin itself has not shown consistent efficacy in major depression.[206,207] This disappointing result has increased interest in the use of agomelatine, a melatonin receptor agonist and 5-hydroxytryptamine (serotonin) receptor 2C antagonist. Like established antidepressants, chronic agomelatine treatment results in increased hippocampal neurogenesis.[208] In multiple mouse models of anxiety, agomelatine has shown efficacy.[209] Interestingly, although agomelatine was effective at all times of day when administered alone, when it was given concomitantly with a melatonin receptor antagonist, agomelatine was effective only when given in the morning.[210] This finding suggests that agomelatine produces its anxiolytic/antidepressant effects through both melatonergic and serotonergic mechanisms.

Loo and colleagues[211] conducted a double-blind, placebo-controlled trial of agomelatine involving 711 patients with MDD and found that 25 mg of agomelatine was significantly more effective than placebo. Furthermore, agomelatine separated from placebo at 2 weeks of treatment, and this suggested that the medication may have a more rapid onset of action than existing medications. A second double-blind, placebo-controlled trial confirmed the efficacy of agomelatine, and a pooled analysis confirmed the overall effect.[212] Two additional trials found agomelatine to be as effective as venlafaxine in the treatment of MDD.[213,214] Together, these findings have raised substantial hope that agomelatine will prove to be an effective antidepressant, possibly with a faster onset of action than current antidepressants. It is important to note that the long-term efficacy and tolerability of agomelatine have not yet been reported.

Omega-3 Fatty Acids

Omega-3 fatty acids are essential fatty acids found most abundantly in seafood. Most epidemiologic studies show an inverse correlation between consumption of omega-3 fatty acids and depression rates.[215] Furthermore, depressed patients have lower levels of omega-3 fatty acids than controls.[216] Omega-3 fatty acids have been tested in depression in at least 10 placebo-controlled trials with mixed results. In a meta-analysis, Lin and Su[217] found significant antidepressant efficacy for omega-3 fatty acids but also found evidence for heterogeneity and publication bias. Thus, further large, well-designed studies are necessary to definitively determine the efficacy of omega-3 fatty acids in depression.

Cholinergic Agents

The cholinergic system has been of peripheral interest to the study of mood disorders for over 25 years since Janowsky's seminal work in the field[218] (see Picciotto *et al.*[219] for a review). Several studies have identified abnormalities of the cholinergic system associated with MDD.[220] More recently, it has also been recognized that many of the currently approved antidepressant agents also antagonize nicotinic acetylcholine receptors at clinically effective doses.[221] These findings, in addition to the established association between cigarette smoking, nicotine withdrawal, and depression,[222,223] led to a renewed interest in pursuing nicotinic agents as antidepressant treatments. Recent preclinical studies have demonstrated antidepressant-like effects of nicotinic antagonists and partial agonists in rodent models of depression.[224,225] In humans, a small clinical trial using the nicotinic antagonist mecamylamine in combination with SSRIs suggested that the drug has efficacy in the treatment of depression (T. George, unpublished data, 2008). A 2006 press release by Targacept, Inc., also reported the favorable results of a phase II clinical trial with TRIDMAC, a treatment combination consisting of mecamylamine hydrochloride as an augmentation treatment to citalopram hydrobromide; however, the study data remain unpublished.

There has also been an interesting and unexpected finding suggesting that the intravenous infusion of scopolamine, a muscarinic antagonist, is associated with a rapid and robust antidepressant response.[226] Patients with MDD and bipolar disorder experienced improvement in their mood as early as the evening of the infusion and continued to improve after additional infusions. The effect in this small study seemed to last for several days to weeks, and this suggested that the effect was not simply related to the euphoric effect that can be associated with anticholinergic drugs. However, similarly to the case for NMDA receptor antagonists, more work is required to better characterize the true antidepressant potential of anticholinergic drugs and to determine the actual mechanism of the antidepressant response (Table 1).

CONCLUSIONS

Although existing antidepressants are effective for a proportion of patients, it is increasingly clear that there are substantial limitations to our current armamentarium. Most patients treated with first-line antidepressants do not achieve remission, and further adjustments to pharmacotherapy have only marginal benefits.[227] In addition, even when effective, current antidepressants take 2 to 5 weeks to work[228] and leave patients at a high risk of relapse.[6] Despite these limitations, there have not been major improvements in pharmacotherapy for depression since the development of SSRIs 2 decades ago. This is in large part due to the difficulties in developing new medications through rational drug design in the absence of a pathophysiological model. With advances in our understanding of the neurobiological complexity underlying depression, however, there is reason to be hopeful that a new generation of antidepressants could help us truly improve our treatments. Perhaps the greatest strengths of the emerging treatments are the breadth of neurobiological systems targeted and the number of new medications based on emerging pathophysiological models. These various treatments may affect our management of depression in different ways and may help to personalize treatment by targeting specific systems with demonstrated abnormalities in the individual.

Although triple reuptake inhibitors, atypical antipsychotics, and agomelatine are likely to be available as antidepressants in the near future, clinical studies to date suggest that these drugs provide only an incremental improvement in efficacy in comparison with existing treatments. In contrast, treatments targeting amino acid neurotransmitter, immune, and HPA systems may provide qualitative advances in our ability to treat depression, especially in subjects with proven resistance to the current medications. In general, drugs targeting these systems require more development time before they are available for regular clinical use, although some compounds such as riluzole, NMDA antagonists, TNFα antagonists, RU486, and metyrapone are currently in clinical use and could rapidly be repositioned as antidepressant agents.

Another important facet of having a larger arsenal of drugs targeting multiple neurobiological systems is that it provides a greater opportunity to selectively target specific symptom complexes and potentially subgroups of patients sharing common pathophysiological profiles. Having many classes of available antidepressants and antidepressants that target many different systems is of increasing importance as we enter the era of pharmacogenomics and personalized medicine.

ACKNOWLEDGMENT

This study was supported by grant K02 MH076222-01A1 (to Gerard Sanacora), the Donaghue Foundation Award (to Srijan Sen), the Daniel X. and Mary Freedman Foundation, the VA CT Research Enhancement Award Program Research Center (to Gerard Sanacora and Srijan Sen), and the National Alliance for Research on Schizophrenia and Depression Independent Investigator Award (to Gerard Sanacora).

Gerard Sanacora is a cosponsor of a filed patent application (PCTWO06108055A1) related to the use of drugs that modulate glutamate neurotransmission for the treatment of depression.

DISCLOSURES

Dr. Gerard Sanacora has received consulting and/or research funds from Abbott, AstraZeneca, Bristol-Myers Squibb, Eli Lilly, Lundbeck, Pfizer, Roche, Ruxton, and Sepracor. He is a co-sponsor of a filed patent application by Yale University (PCTWO06108055A1) related to the use of glutamate-modulating drugs for the treatment of depression.

Table 1. *Summary of Evidence Specifically Supporting Neurotransmitter Systems as Targets for Antidepressants.*

Target System	Preclinical Evidence	Clinical Evidence	Most Advanced Clinical Phase
Neurotrophic factors	Strong evidence of neurotrophic factor involvement in pathophysiology and antidepressant mechanism of action; good evidence for a direct antidepressant effect in rodents	None beyond the indirect effects of rolipram	No clinical studies to date
Glutamate	Strong evidence with various targets in several studies with several rodent models	Evidence of acute antidepressant effects in small randomized clinical trials	Early phase clinical trials with NMDA antagonist and AMPA agents ongoing
GABA	Modest evidence in a few rodent studies	Benzodiazepines as a whole not effective; some emerging data for another class of GABAergic drugs	Early phase clinical trials with other GABAergic agents ongoing
HPA	Strong evidence with various targets in several studies employing several rodent models	Preliminary data on stress responsivity now emerging	Several large CRF antagonist trials underway
Immune modulating agents	Strong evidence that immune modulating agents can produce depressive-like behaviors; modest evidence of antidepressant activity	Very limited data from patients being treated for autoimmune disorders	Several small early phase trials underway
Neurokinins	Strong evidence implicating the neurokinins in the pathophysiology of depression and anxiety	Somewhat disappointing results from a small number of randomized, placebo-controlled trials	Trials of NK-1 antagonists appear halted, but some continued interest in other targets in the neurokinin system
Monoaminergic	Strongest evidence of antidepressant-like effects in animal models	Established antidepressant effects	Newer agents including triple uptake inhibitors in clinical trials
Cholinergic	Strong evidence of cholinergic contributions to the pathophysiology of MDD; less evidence for antidepressant action	Limited small clinical trails that appear positive for nicotinic antagonists; single study with a muscarinic antagonist	Nicotinic drugs in later phase clinical development

Abbreviations: AMPA, α-amino-3-hydroxy-5-methyl-isoxazole-4-propionic acid; CRF, corticotropin-releasing factor; GABA, gamma aminobutyric acid; HPA, hypothalamic-pituitary-adrenal; MDD, major depressive disorder; NK-1, neurokinin-1; NMDA, *N*-methyl-D-aspartate.

REFERENCES

1. Kessler RC, Berglund P, Demler O, et al. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 2005; 62: 593–602.
2. Murray CJ, Lopez AD. Alternative projections of mortality and disability by cause 1990–2020: global burden of disease study. *Lancet* 1997; 349: 1498–1504.
3. Jans LA, Riedel WJ, Markus CR, Blokland A. Serotonergic vulnerability and depression: assumptions,

experimental evidence and implications. *Mol Psychiatry* 2007; 12: 522–543.

4. Ruhe HG, Mason NS, Schene AH. Mood is indirectly related to serotonin, norepinephrine and dopamine levels in humans: a meta-analysis of monoamine depletion studies. *Mol Psychiatry* 2007; 12: 331–359.
5. Bhagwagar Z, Hinz R, Taylor M, et al. Increased 5-HT(2A) receptor binding in euthymic, medication-free patients recovered from depression: a positron emission study with [(11)C]MDL 100,907. *Am J Psychiatry* 2006; 163: 1580–1587.
6. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry* 2006; 163: 1905–1917.
7. Gould E, Reeves AJ, Graziano MS, Gross CG. Neurogenesis in the neocortex of adult primates. *Science* 1999; 286: 548–552.
8. Duman RS, Monteggia LM. A neurotrophic model for stress-related mood disorders. *Biol Psychiatry* 2006; 59: 1116–1127.
9. Altar CA, Cai N, Bliven T, et al. Anterograde transport of brain-derived neurotrophic factor and its role in the brain. *Nature* 1997; 389: 856–860.
10. Dwivedi Y, Rizavi HS, Conley RR, et al. Altered gene expression of brain-derived neurotrophic factor and receptor tyrosine kinase B in postmortem brain of suicide subjects. *Arch Gen Psychiatry* 2003; 60: 804–815.
11. Bannerman DM, Lemaire M, Yee BK, et al. Selective cytotoxic lesions of the retrohippocampal region produce a mild deficit in social recognition memory. *Exp Brain Res* 2002; 142: 395–401.
12. Videbech P, Ravnkilde B. Hippocampal volume and depression: a meta-analysis of MRI studies. *Am J Psychiatry* 2004; 161: 1957–1966.
13. Heldt SA, Stanek L, Chhatwal JP, Ressler KJ. Hippocampus-specific deletion of BDNF in adult mice impairs spatial memory and extinction of aversive memories. *Mol Psychiatry* 2007; 12: 656–670.
14. Sklar P, Gabriel SB, McInnis MG, et al. Family-based association study of 76 candidate genes in bipolar disorder: BDNF is a potential risk locus. Brain-derived neutrophic factor. *Mol Psychiatry* 2002; 7: 579–593.
15. Sen S, Nesse RM, Stoltenberg SF, et al. A BDNF coding variant is associated with the NEO personality inventory domain neuroticism, a risk factor for depression. *Neuropsychopharmacology* 2003; 28: 397–401.
16. Kaufman J, Yang BZ, Douglas-Palumberi H, et al. Brain-derived neurotrophic factor-5-HTTLPR gene interactions and environmental modifiers of depression in children. *Biol Psychiatry* 2006; 59: 673–680.
17. Siuciak JA, Lewis DR, Wiegand SJ, Lindsay RM. Antidepressant-like effect of brain-derived neurotrophic factor (BDNF). *Pharmacol Biochem Behav* 1997; 56: 131–137.
18. Tsankova NM, Berton O, Renthal W, et al. Sustained hippocampal chromatin regulation in a mouse model of depression and antidepressant action. *Nat Neurosci* 2006; 9: 519–525.
19. Tanis KQ, Newton SS, Duman RS. Targeting neurotrophic/growth factor expression and signaling for antidepressant drug development. *CNS Neurol Disord Drug Targets* 2007; 6: 151–160.
20. Warner-Schmidt JL, Duman RS. Hippocampal neurogenesis: opposing effects of stress and antidepressant treatment. *Hippocampus* 2006; 16: 239–249.
21. Sairanen M, Lucas G, Ernfors P, et al. Brain-derived neurotrophic factor and antidepressant drugs have different but coordinated effects on neuronal turnover, proliferation, and survival in the adult dentate gyrus. *J Neurosci* 2005; 25: 1089–1094.
22. Monteggia LM, Barrot M, Powell CM, et al. Essential role of brain-derived neurotrophic factor in adult hippocampal function. *Proc Natl Acad Sci U S A* 2004; 101: 10,827–10,832.
23. Berton O, McClung CA, Dileone RJ, et al. Essential role of BDNF in the mesolimbic dopamine pathway in social defeat stress. *Science* 2006; 311: 864–868.
24. Newton SS, Girgenti MJ, Collier EF, Duman RS. Electroconvulsive seizure increases adult hippocampal angiogenesis in rats. *Eur J Neurosci* 2006; 24: 819–828.
25. Khawaja X, Xu J, Liang JJ, Barrett JE. Proteomic analysis of protein changes developing in rat hippocampus after chronic antidepressant treatment: implications for depressive disorders and future therapies. *J Neurosci Res* 2004; 75: 451–460.
26. Evans SJ, Choudary PV, Neal CR, et al. Dysregulation of the fibroblast growth factor system in major depression. *Proc Natl Acad Sci U S A* 2004; 101: 15,506–15,511.
27. Pollack SJ, Harper SJ. Small molecule Trk receptor agonists and other neurotrophic factor mimetics. *Curr Drug Targets CNS Neurol Disord* 2002; 1: 59–80.
28. Fleischhacker WW, Hinterhuber H, Bauer H, et al. A multicenter double-blind study of three different doses of the new cAMP-phosphodiesterase inhibitor rolipram in patients with major depressive disorder. *Neuropsychobiology* 1992; 26: 59–64.
29. Fujimaki K, Morinobu S, Duman RS. Administration of a cAMP phosphodiesterase 4 inhibitor enhances antidepressant-induction of BDNF mRNA in rat hippocampus. *Neuropsychopharmacology* 2000; 22: 42–51.
30. Jeon YH, Heo YS, Kim CM, et al. Phosphodiesterase: overview of protein structures, potential therapeutic applications and recent progress in drug development. *Cell Mol Life Sci* 2005; 62: 1198–1220.
31. Gould TD. Targeting glycogen synthase kinase-3 as an approach to develop novel mood-stabilising medications. *Expert Opin Ther Targets* 2006; 10: 377–392.
32. Kugaya A, Sanacora G. Beyond monoamines: glutamatergic function in mood disorders. *CNS Spectr* 2005; 10: 808–819.
33. Krystal JH, Sanacora G, Blumberg H, et al. Glutamate and GABA systems as targets for novel antidepressant and mood-stabilizing treatments. *Mol Psychiatry* 2002; 7(suppl 1): S71–S80.
34. Pittaluga A, Raiteri L, Longordo F, et al. Antidepressant treatments and function of glutamate ionotropic receptors mediating amine release in hippocampus. *Neuropharmacology* 2007; 53: 27–36.
35. Moghaddam B, Jackson M. Effect of stress on prefrontal cortex function. *Neurotox Res* 2004; 6: 73–78.

36. Sapolsky RM. The possibility of neurotoxicity in the hippocampus in major depression: a primer on neuron death. *Biol Psychiatry* 2000; 48: 755–765.
37. Czeh B. Simon M, Schmelting B, et al. Astroglial plasticity in the hippocampus is affected by chronic psychosocial stress and concomitant fluoxetine treatment. *Neuropsychopharmacology* 2006; 31: 1616–1626.
38. Czeh B, Muller-Keuker JI, Rygula R, et al. Chronic social stress inhibits cell proliferation in the adult medial prefrontal cortex: hemispheric asymmetry and reversal by fluoxetine treatment. *Neuropsychopharmacology* 2007; 32: 1490–1503.
39. Rajkowska G. Miguel-Hidalgo JJ. Gliogenesis and glial pathology in depression. *CNS Neurol Disord Drug Targets* 2007; 6: 219–233.
40. Nowak G. Ordway GA, Paul IA. Alterations in the N-methyl-D-aspartate (NMDA) receptor complex in the frontal cortex of suicide victims. *Brain Res* 1995; 675: 157–164.
41. Nudmamud S, Reynolds LM, Reynolds GP. N-acetylaspartate and N-acetylaspartylglutamate deficits in superior temporal cortex in schizophrenia and bipolar disorder: a postmortem study. *Biol Psychiatry* 2003; 53: 1138–1141.
42. Karolewicz B, Stockmeier CA, Ordway GA. Elevated levels of the NR2C subunit of the NMDA receptor in the locus coeruleus in depression. *Neuropsychopharmacology* 2005; 30: 1557–1567.
43. Kristiansen LV, Meador-Woodruff JH. Abnormal striatal expression of transcripts encoding NMDA interacting PSD proteins in schizophrenia, bipolar disorder and major depression. *Schizophr Res* 2005; 78: 87–93.
44. McCullumsmith RE, Kristiansen LV, Beneyto M, et al. Decreased NR1, NR2A, and SAP102 transcript expression in the hippocampus in bipolar disorder. *Brain Res* 2007; 1127: 108–118.
45. Bhagwagar Z, Wylezinska M, Jezzard P, et al. Reduction in occipital cortex gamma-aminobutyric acid concentrations in medication-free recovered unipolar depressed and bipolar subjects. *Biol Psychiatry* 2007; 15:61: 806–812.
46. Hasler G, van der Veen JW, Tumonis T, et al. Reduced prefrontal glutamate/glutamine and gamma-aminobutyric acid levels in major depression determined using proton magnetic resonance spectroscopy. *Arch Gen Psychiatry* 2007; 64: 193–200.
47. Sanacora G, Gueorguieva R, Epperson CN, et al. Subtype-specific alterations of GABA and glutamate in major depression. *Arch Gen Psychiatry* 2004; 61: 705–713.
48. Michael N, Erfurth A, Ohrmann P, et al. Metabolic changes within the left dorsolateral prefrontal cortex occurring with electroconvulsive therapy in patients with treatment resistant unipolar depression. *Psychol Med* 2003; 33: 1277–1284.
49. Pfleiderer B, Michael N, Erfurth A, et al. Effective electroconvulsive therapy reverses glutamate/glutamine deficit in the left anterior cingulum of unipolar depressed patients. *Psychiatry Res* 2003; 122: 185–192.
50. Auer DP, Putz B, Kraft E, et al. Reduced glutamate in the anterior cingulate cortex in depression: an in vivo proton magnetic resonance spectroscopy study. *Biol Psychiatry* 2000; 47: 305–313.
51. Hardingham GE. Pro-survival signalling from the NMDA receptor. *Biochem Soc Trans* 2006;34:pt 5; 936–938.
52. Hardingham GE, Bading H. The yin and yang of NMDA receptor signalling. *Trends Neurosci* 2003; 26: 81–89.
53. Lau CG, Zukin RS. NMDA receptor trafficking in synaptic plasticity and neuropsychiatric disorders. *Nat Rev Neurosci* 2007; 8: 413–426.
54. Skolnick P. Antidepressants for the new millennium. *Eur J Pharmacol* 1999; 375: 31–40.
55. Paul IA, Nowak G, Layer RT, et al. Adaption of N-methyl-D-aspartate receptor complex following antidepressant treatments. *J Pharmacol Exp Ther* 1994; 269: 95–102.
56. Berman RM, Cappiello A, Anand A, et al. Antidepressant effects of ketamine in depressed patients. *Biol Psychiatry* 2000; 47: 351–354.
57. Zarate CA Jr, Singh JB, Carlson PJ, et al. A randomized trial of an N-methyl-D-aspartate antagonist in treatment-resistant major depression. *Arch Gen Psychiatry* 2006; 63: 856–864.
58. Carr DB, Goudas LC, Denman WT, et al. Safety and efficacy of intranasal ketamine in a mixed population with chronic pain. *Pain* 2004; 110: 762–764.
59. Olney JW. Neurotoxicity of NMDA receptor antagonists: an overview. *Psychopharmacol Bull* 1994; 30: 533–540.
60. Soriano FX. Hardingham GE. Compartmentalized NMDA receptor signalling to survival and death. *J Physiol* 2007;584:pt 2; 381–387.
61. Preskorn S, Baker B, Omo K. et al. *A placebo-controlled trial of the NR2B specific NMDA antagonist CP-101,606 plus paroxetine for treatment resistant depression (TRD).* San Diego, CA: American Psychiatric Association; 2007; 154 NR362.
62. Malinow R. Malenka RC. AMPA receptor trafficking and synaptic plasticity. *Annu Rev Neurosci* 2002; 25: 103–126.
63. Zhu JJ, Qin Y, Zhao M, et al. Ras and Rap control AMPA receptor trafficking during synaptic plasticity. *Cell* 2002; 110: 443–455.
64. Jaskolski F, Coussen F, Mulle C. Subcellular localization and trafficking of kainate receptors. *Trends Pharmacol Sci* 2005; 26: 20–26.
65. Lerma J. Kainate receptor physiology. *Curr Opin Pharmacol* 2006; 6: 89–97.
66. Bleakman D, Alt A, Witkin JM. AMPA receptors in the therapeutic management of depression. *CNS Neurol Disord Drug Targets* 2007; 6: 117–126.
67. Du J, Suzuki K, Wei Y, et al. The anticonvulsants lamotrigine, riluzole, and valproate differentially regulate AMPA receptor membrane localization: relationship to clinical effects in mood disorders. *Neuropsychopharmacology* 2007; 32: 793–802.
68. Maeng S, Zarate CA Jr, Du J, et al. Cellular mechanisms underlying the antidepressant effects of ketamine: role of alpha-amino-3-hydroxy-5-methylisoxazole-4-propionic acid receptors. *Biol Psychiatry* 2008; 63: 349–352.
69. Paddock S, Laje G, Charney D, et al. Association of GRIK4 with outcome of antidepressant treatment in the STAR*D cohort. *Am J Psychiatry* 2007; 164: 1181–1188.
70. Laje G, Paddock S, Manji H, et al. Genetic markers of suicidal ideation emerging during citalopram

treatment of major depression. *Am J Psychiatry* 2007; 164: 1530–1538.

71. Alt A, Nisenbaum ES, Bleakman D, Witkin JM. A role for AMPA receptors in mood disorders. *Biochem Pharmacol* 2006; 71: 1273–1288.
72. O'Neill MJ, Witkin JM. AMPA receptor potentiators: application for depression and Parkinson's disease. *Curr Drug Targets* 2007; 8: 603–620.
73. Conn PJ. Physiological roles and therapeutic potential of metabotropic glutamate receptors. *Ann N Y Acad Sci* 2003; 1003: 12–21.
74. D'Ascenzo M, Fellin T, Terunuma M, et al. mGluR5 stimulates gliotransmission in the nucleus accumbens. *Proc Natl Acad Sci U S A* 2007; 104: 1995–2000.
75. Palucha A, Pilc A. Metabotropic glutamate receptor ligands as possible anxiolytic and antidepressant drugs. *Pharmacol Ther* 2007; 115: 116–147.
76. Witkin JM, Marek GJ, Johnson BG, Schoepp DD. Metabotropic glutamate receptors in the control of mood disorders. *CNS Neurol Disord Drug Targets* 2007; 6: 87–100.
77. Corti C, Battaglia G, Molinaro G, et al. The use of knock-out mice unravels distinct roles for mGlu2 and mGlu3 metabotropic glutamate receptors in mechanisms of neurodegeneration/neuroprotection. *J Neurosci* 2007; 27: 8297–8308.
78. Chaki S, Yoshikawa R, Hirota S, et al. MGS0039: a potent and selective group II metabotropic glutamate receptor antagonist with antidepressant-like activity. *Neuropharmacology* 2004; 46: 457–467.
79. Karasawa J, Shimazaki T, Kawashima N, Chaki S. AMPA receptor stimulation mediates the antidepressant-like effect of a group II metabotropic glutamate receptor antagonist. *Brain Res* 2005; 1042: 92–98.
80. Dunayevich E, Erickson J, Levine L, et al. Efficacy and tolerability of an mGlu2/3 agonist in the treatment of generalized anxiety disorder. *Neuropsychopharmacology* 2007: doi://10.1038/sj.npp.1301531. Available at: http://www.nature.com.
81. Patil ST, Zhang L, Martenyi F, et al. Activation of mGlu2/3 receptors as a new approach to treat schizophrenia: a randomized phase 2 clinical trial. *Nat Med* 2007; 13: 1102–1107.
82. Miller RG, Mitchell JD, Lyon M, Moore DH. Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). *Cochrane Database Syst Rev* 2007; CD001447.
83. Benavides J, Camelin JC, Mitrani N, et al. 2-Amino-6-trifluoromethoxy benzothiazole, a possible antagonist of excitatory amino acid neurotransmission–II. Biochemical properties. *Neuropharmacology* 1985; 24: 1085–1092.
84. Mizoule J, Meldrum B, Mazadier M, et al. 2-Amino-6-trifluoromethoxy benzothiazole, a possible antagonist of excitatory amino acid neurotransmission–I. Anticonvulsant properties. *Neuropharmacology* 1985; 24: 767–773.
85. Debono MW, Le Guern J, Canton T, et al. Inhibition by riluzole of electrophysiological responses mediated by rat kainate and NMDA receptors expressed in Xenopus oocytes. *Eur J Pharmacol* 1993; 235: 283–289.
86. Benoit E, Escande D. Riluzole specifically blocks inactivated Na channels in myelinated nerve fibre. *Pflugers Arch* 1991; 419: 603–609.
87. Hebert T, Drapeau P, Pradier L, Dunn RJ. Block of the rat brain IIA sodium channel alpha subunit by the neuroprotective drug riluzole. *Mol Pharmacol* 1994; 45: 1055–1060.
88. Zona C, Siniscalchi A, Mercuri NB, Bernardi G. Riluzole interacts with voltage-activated sodium and potassium currents in cultured rat cortical neurons. *Neuroscience* 1998; 85: 931–938.
89. Urbani A, Belluzzi O. Riluzole inhibits the persistent sodium current in mammalian CNS neurons. *Eur J Neurosci* 2000; 12: 3567–3574.
90. Huang CS, Song JH, Nagata K, et al. Effects of the neuroprotective agent riluzole on the high voltage-activated calcium channels of rat dorsal root ganglion neurons. *J Pharmacol Exp Ther* 1997; 282: 1280–1290.
91. Stefani A, Spadoni F, Bernardi G. Differential inhibition by riluzole, lamotrigine, and phenytoin of sodium and calcium currents in cortical neurons: implications for neuroprotective strategies. *Exp Neurol* 1997; 147: 115–122.
92. Duprat F, Lesage F, Patel AJ, et al. The neuroprotective agent riluzole activates the two P domain K(+) channels TREK-1 and TRAAK. *Mol Pharmacol* 2000; 57: 906–912.
93. Xu L, Enyeart JA, Enyeart JJ. Neuroprotective agent riluzole dramatically slows inactivation of Kv1.4 potassium channels by a voltage-dependent oxidative mechanism. *J Pharmacol Exp Ther* 2001; 299: 227–237.
94. Ahn HS, Choi JS, Choi BH, et al. Inhibition of the cloned delayed rectifier K+ channels, Kv1.5 and Kv3.1, by riluzole. *Neuroscience* 2005; 133: 1007–1019.
95. Ahn HS, Kim SE, Jang HJ, et al. Interaction of riluzole with the closed inactivated state of Kv4.3 channels. *J Pharmacol Exp Ther* 2006; 319: 323–331.
96. Cheramy A, Barbeito L, Godeheu G, Glowinski J. Riluzole inhibits the release of glutamate in the caudate nucleus of the cat in vivo. *Neurosci Lett* 1992; 147: 209–212.
97. Doble A, Hubert JP, Blanchard JC. Pertussis toxin pretreatment abolishes the inhibitory effect of riluzole and carbachol on D-[3H]aspartate release from cultured cerebellar granule cells. *Neurosci Lett* 1992; 140: 251–254.
98. Martin D, Thompson MA, Nadler JV. The neuroprotective agent riluzole inhibits release of glutamate and aspartate from slices of hippocampal area CA1. *Eur J Pharmacol* 1993; 250: 473–476.
99. Jehle T, Bauer J, Blauth E, et al. Effects of riluzole on electrically evoked neurotransmitter release. *Br J Pharmacol* 2000; 130: 1227–1234.
100. Azbill RD, Mu X, Springer JE. Riluzole increases high-affinity glutamate uptake in rat spinal cord synaptosomes. *Brain Res* 2000; 871: 175–180.
101. Dunlop J, Beal McIlvain H, She Y, Howland DS. Impaired spinal cord glutamate transport capacity and reduced sensitivity to riluzole in a transgenic superoxide dismutase mutant rat model of amyotrophic lateral sclerosis. *J Neurosci* 2003; 23: 1688–1696.
102. Sung B, Lim G, Mao J. Altered expression and uptake activity of spinal glutamate transporters after nerve injury contribute to the pathogenesis of neuropathic pain in rats. *J Neurosci* 2003; 23: 2899–2910.

103. Katoh-Semba R, Asano T, Ueda H, et al. Riluzole enhances expression of brain-derived neurotrophic factor with consequent proliferation of granule precursor cells in the rat hippocampus. *FASEB J* 2002; 16: 1328–1330.
104. Coric V, Milanovic S, Wasylink S, et al. Beneficial effects of the antiglutamatergic agent riluzole in a patient diagnosed with obsessive-compulsive disorder and major depressive disorder. *Psychopharmacology* 2003; 167: 219–220.
105. Zarate CA Jr, Payne JL, Quiroz J, et al. An open-label trial of riluzole in patients with treatment-resistant major depression. *Am J Psychiatry* 2004; 161: 171–174.
106. Zarate CA Jr, Quiroz JA. Singh JB, et al. An open-label trial of the glutamate-modulating agent riluzole in combination with lithium for the treatment of bipolar depression. *Biol Psychiatry* 2005; 57: 430–432.
107. Sanacora G, Kendell SF, Levin Y, et al. Preliminary evidence of riluzole efficacy in antidepressant-treated patients with residual depressive symptoms. *Biol Psychiatry* 2007; 61: 822–825.
108. Mathew SJ, Amiel JM, Coplan JD, et al. Open-label trial of riluzole in generalized anxiety disorder. *Am J Psychiatry* 2005; 162: 2379–2381.
109. Sanacora G, Saricicek A. GABAergic contributions to the pathophysiology of depression and the mechanism of antidepressant action. *CNS Neurol Disord Drug Targets* 2007; 6: 127–140.
110. Schatzberg AF, Cole JO. Benzodiazepines in depressive disorders. *Arch Gen Psychiatry* 1978; 35: 1359–1365.
111. Bateson AN. The benzodiazepine site of the GABAA receptor: an old target with new potential? *Sleep Med* 2004; 5(suppl 1): S9–S15.
112. Russell MG, Carling RW, Street LJ, et al. Discovery of imidazo[1,2-b][1,2,4]triazines as GABA(A) alpha2/3 subtype selective agonists for the treatment of anxiety. *J Med Chem* 2006; 49: 1235–1238.
113. Dias R, Sheppard WF, Fradley RL, et al. Evidence for a significant role of alpha 3-containing GABAA receptors in mediating the anxiolytic effects of benzodiazepines. *J Neurosci* 2005; 25: 10,682–10,688.
114. Fava M, McCall WV, Krystal A, et al. Eszopiclone co-administered with fluoxetine in patients with insomnia coexisting with major depressive disorder. *Biol Psychiatry* 2006; 59: 1052–1060.
115. Berger M, Krieg C, Bossert S, et al. Past and present strategies of research on the HPA-axis in psychiatry. *Acta Psychiatr Scand Suppl* 1988; 341: 112–125.
116. Ising M, Lauer CJ, Holsboer F, Modell S. The Munich vulnerability study on affective disorders: premorbid neuroendocrine profile of affected high-risk probands. *J Psychiatr Res* 2005; 39: 21–28.
117. Heuser I. Anna-Monika-Prize paper. The hypothalamic-pituitary-adrenal system in depression. *Pharmacopsychiatry* 1998; 31: 10–13.
118. Reul JM, Holsboer F. Corticotropin-releasing factor receptors 1 and 2 in anxiety and depression. *Curr Opin Pharmacol* 2002; 2: 23–33.
119. Stenzel-Poore MP, Heinrichs SC, Rivest S, et al. Overproduction of corticotropin-releasing factor in transgenic mice: a genetic model of anxiogenic behavior. *J Neurosci* 1994;14:pt 1: 2579–2584.
120. Britton KT, Morgan J, Rivier J, et al. Chlordiazepoxide attenuates response suppression induced by corticotropin-releasing factor in the conflict test. *Psychopharmacology (Berl)* 1985; 86: 170–174.
121. Krahn DD, Gosnell BA, Grace M, Levine AS. CRF antagonist partially reverses CRF- and stress-induced effects on feeding. *Brain Res Bull* 1986; 17: 285–289.
122. Nemeroff CB, Bissette G, Akil H, Fink M. Neuropeptide concentrations in the cerebrospinal fluid of depressed patients treated with electroconvulsive therapy. Corticotrophin-releasing factor, beta-endorphin and somatostatin. *Br J Psychiatry* 1991; 158: 59–63.
123. Carpenter RE, Watt MJ, Forster GL, et al. Corticotropin releasing factor induces anxiogenic locomotion in trout and alters serotonergic and dopaminergic activity. *Horm Behav* 2007; 52: 600–611.
124. Gutman DA, Owens MJ, Skelton KH, et al. The corticotropin-releasing factor1 receptor antagonist R121919 attenuates the behavioral and endocrine responses to stress. *J Pharmacol Exp Ther* 2003; 304: 874–880.
125. Holsboer F, Muller OA, Doerr HG, et al. ACTH and multisteroid responses to corticotropin-releasing factor in depressive illness: relationship to multisteroid responses after ACTH stimulation and dexamethasone suppression. *Psychoneuroendocrinology* 1984; 9: 147–160.
126. Britton KT, Lee G, Vale W, et al. Corticotropin releasing factor (CRF) receptor antagonist blocks activating and 'anxiogenic' actions of CRF in the rat. *Brain Res* 1986; 369: 303–306.
127. Raux-Demay MC, Girard F. The physiology of corticotropin-releasing factor (CRF). *Reprod Nutr Dev* 1985; 25: 931–943.
128. Dunn AJ, Berridge CW. Physiological and behavioral responses to corticotropin-releasing factor administration: is CRF a mediator of anxiety or stress responses? *Brain Res Brain Res Rev* 1990; 15: 71–100.
129. Smith GW, Aubry JM, Dellu F, et al. Corticotropin releasing factor receptor 1-deficient mice display decreased anxiety, impaired stress response, and aberrant neuroendocrine development. *Neuron* 1998; 20: 1093–1102.
130. Timpl P, Spanagel R, Sillaber I, et al. Impaired stress response and reduced anxiety in mice lacking a functional corticotropin-releasing hormone receptor 1. *Nat Genet* 1998; 19: 162–166.
131. Skutella T, Probst JC, Renner U, et al. Corticotropin-releasing hormone receptor (type I) antisense targeting reduces anxiety. *Neuroscience* 1998; 85: 795–805.
132. Lovenberg TW, Chalmers DT, Liu C, De Souza EB. CRF2 alpha and CRF2 beta receptor mRNAs are differentially distributed between the rat central nervous system and peripheral tissues. *Endocrinology* 1995; 136: 4139–4142.
133. Bale TL, Contarino A, Smith GW, et al. Mice deficient for corticotropin-releasing hormone receptor-2 display anxiety-like behaviour and are hypersensitive to stress. *Nat Genet* 2000; 24: 410–414.
134. Valdez GR. Development of CRF1 receptor antagonists as antidepressants and anxiolytics: progress to date. *CNS Drugs* 2006; 20: 887–896.
135. Kehne J, De Lombaert S. Non-peptidic CRF1 receptor antagonists for the treatment of anxiety, depression and stress disorders. *Curr Drug Targets CNS Neurol Disord* 2002; 1: 467–493.

136. Chaki S, Nakazato A, Kennis L, et al. Anxiolytic- and antidepressant-like profile of a new CRF1 receptor antagonist, R278995 CRA0450. *Eur J Pharmacol* 2004; 485: 145–158.
137. Jutkiewicz EM, Wood SK, Houshyar H, et al. The effects of CRF antagonists, antalarmin, CP154,526, LWH234, and R121919, in the forced swim test and on swim-induced increases in adrenocorticotropin in rats. *Psychopharmacology (Berl)* 2005; 180: 215–223.
138. Li YW, Fitzgerald L, Wong H, et al. The pharmacology of DMP696 and DMP904, non-peptidergic CRF1 receptor antagonists. *CNS Drug Rev* 2005; 11: 21–52.
139. Overstreet DH, Griebel G. Antidepressant-like effects of CRF1 receptor antagonist SSR125543 in an animal model of depression. *Eur J Pharmacol* 2004; 497: 49–53.
140. Zorrilla EP, Koob GF. The therapeutic potential of CRF1 antagonists for anxiety. *Expert Opin Investig Drugs* 2004; 13: 799–828.
141. Zobel AW, Nickel T, Kunzel HE, et al. Effects of the high-affinity corticotropin-releasing hormone receptor 1 antagonist R121919 in major depression: the first 20 patients treated. *J Psychiatr Res* 2000; 34: 171–181.
142. Ising M, Zimmermann US, Kunzel HE, et al. High-affinity CRF1 receptor antagonist NBI-34,041: preclinical and clinical data suggest safety and efficacy in attenuating elevated stress response. *Neuropsychopharmacology* 2007; 32: 1941–1949.
143. Whitnall MH, Gainer H. Ultrastructural immunolocalization of vasopressin and neurophysin in neurosecretory cells of dehydrated rats. *Brain Res* 1985; 361: 400–404.
144. Bhattacharya SK, Chakrabarti A, Glover V. Stress and water balance: the roles of ANP, AVP and isatin. *Indian J Exp Biol* 1998; 36: 1195–1200.
145. Landgraf R, Gerstberger R, Montkowski A, et al. V1 vasopressin receptor antisense oligodeoxynucleotide into septum reduces vasopressin binding, social discrimination abilities, and anxiety-related behavior in rats. *J Neurosci* 1995; 15: 4250–4258.
146. Kjaer A, Knigge U, Bach FW, Warberg J. Impaired histamine- and stress-induced secretion of ACTH and beta-endorphin in vasopressin-deficient Brattleboro rats. *Neuroendocrinology* 1993; 57: 1035–1041.
147. Egashira N, Tanoue A, Higashihara F, et al. Disruption of the prepulse inhibition of the startle reflex in vasopressin V1b receptor knockout mice: reversal by antipsychotic drugs. *Neuropsychopharmacology* 2005; 30: 1996–2005.
148. Griebel G, Simiand J, Serradeil-Le Gal C, et al. Anxiolytic- and antidepressant-like effects of the non-peptide vasopressin V1b receptor antagonist, SSR149415, suggest an innovative approach for the treatment of stress-related disorders. *Proc Natl Acad Sci U S A* 2002; 99: 6370–6375.
149. Alonso R, Griebel G, Pavone G, et al. Blockade of CRF(1) or V(1b) receptors reverses stress-induced suppression of neurogenesis in a mouse model of depression. *Mol Psychiatry* 2004; 9: 278–286, 24.
150. Dempster EL, Burcescu I, Wigg K, et al. Evidence of an association between the vasopressin V1b receptor gene (AVPR1B) and childhood-onset mood disorders. *Arch Gen Psychiatry* 2007; 64: 1189–1195.
151. Purba JS, Hoogendijk WJ, Hofman MA, Swaab DF. Increased number of vasopressin- and oxytocin-expressing neurons in the paraventricular nucleus of the hypothalamus in depression. *Arch Gen Psychiatry* 1996; 53: 137–143.
152. Rogoz Z, Legutko B. Combined treatment with imipramine and metyrapone induces hippocampal and cortical brain-derived neurotrophic factor gene expression in rats. *Pharmacol Rep* 2005; 57: 840–844.
153. Malison RT, Anand A, Pelton GH, et al. Limited efficacy of ketoconazole in treatment-refractory major depression. *J Clin Psychopharmacol* 1999; 19: 466–470.
154. O'Dwyer AM, Lightman SL, Marks MN, Checkley SA. Treatment of major depression with metyrapone and hydrocortisone. *J Affect Disord* 1995; 33: 123–128.
155. Jahn H, Schick M, Kiefer F, et al. Metyrapone as additive treatment in major depression: a double-blind and placebo-controlled trial. *Arch Gen Psychiatry* 2004; 61: 1235–1244.
156. Young AH, Gallagher P, Watson S, et al. Improvements in neurocognitive function and mood following adjunctive treatment with mifepristone (RU-486) in bipolar disorder. *Neuropsychopharmacology* 2004; 29: 1538–1545.
157. Gaillard RC, Riondel A, Muller AF, et al. RU 486: a steroid with antiglucocorticosteroid activity that only disinhibits the human pituitary-adrenal system at a specific time of day. *Proc Natl Acad Sci U S A* 1984; 81: 3879–3882.
158. Belanoff JK, Rothschild AJ, Cassidy F, et al. An open label trial of C-1073 (mifepristone) for psychotic major depression. *Biol Psychiatry* 2002; 52: 386–392.
159. DeBattista C, Belanoff J, Glass S, et al. Mifepristone versus placebo in the treatment of psychosis in patients with psychotic major depression. *Biol Psychiatry* 2006; 60: 1343–1349.
160. Nihalani ND, Schwartz TL. Mifepristone, a glucocorticoid antagonist for the potential treatment of psychotic major depression. *Curr Opin Investig Drugs* 2007; 8: 563–569.
161. Elenkov IJ, Webster EL, Torpy DJ, Chrousos GP. Stress, corticotropin-releasing hormone, glucocorticoids, and the immune/inflammatory response: acute and chronic effects. *Ann N Y Acad Sci* 1999; 876: 1–11; discussion 11–13.
162. Taylor DL, Jones F, Kubota ES, Pocock JM. Stimulation of microglial metabotropic glutamate receptor mGlu2 triggers tumor necrosis factor alpha-induced neurotoxicity in concert with microglial-derived Fas ligand. *J Neurosci* 2005; 25: 2952–2964.
163. Zou JY, Crews FT. TNF alpha potentiates glutamate neurotoxicity by inhibiting glutamate uptake in organotypic brain slice cultures: neuroprotection by NF kappa B inhibition. *Brain Res* 2005; 1034: 11–24.
164. Chaparro-Huerta V, Rivera-Cervantes MC, Flores-Soto ME, et al. Proinflammatory cytokines and apoptosis following glutamate-induced excitotoxicity mediated by p38 MAPK in the hippocampus of neonatal rats. *J Neuroimmunol* 2005; 165: 53–62.
165. Vesce S, Rossi D, Brambilla L, Volterra A. Glutamate release from astrocytes in physiological conditions and in neurodegenerative disorders characterized by neuroinflammation. *Int Rev Neurobiol* 2007; 82: 57–71.

166. Dantzer R. Cytokine-induced sickness behavior: mechanisms and implications. *Ann N Y Acad Sci* 2001; 933: 222–234.
167. O'Brien SM, Scully P, Fitzgerald P, et al. Plasma cytokine profiles in depressed patients who fail to respond to selective serotonin reuptake inhibitor therapy. *J Psychiatr Res* 2007; 41: 326–331.
168. Fitzgerald P, O'Brien SM, Scully P, et al. Cutaneous glucocorticoid receptor sensitivity and pro-inflammatory cytokine levels in antidepressant-resistant depression. *Psychol Med* 2006; 36: 37–43.
169. Kahl KG, Bens S, Ziegler K, et al. Cortisol, the cortisol-dehydroepiandrosterone ratio, and pro-inflammatory cytokines in patients with current major depressive disorder comorbid with borderline personality disorder. *Biol Psychiatry* 2006; 59: 667–671.
170. Penninx BW, Kritchevsky SB, Yaffe K, et al. Inflammatory markers and depressed mood in older persons: results from the Health, Aging and Body Composition Study. *Biol Psychiatry* 2003; 54: 566–572.
171. Narita K, Murata T, Takahashi T, et al. Plasma levels of adiponectin and tumor necrosis factor-alpha in patients with remitted major depression receiving long-term maintenance antidepressant therapy. *Prog Neuropsychopharmacol Biol Psychiatry* 2006; 30: 1159–1162.
172. Lanquillon S, Krieg JC, Bening-Abu-Shach U, Vedder H. Cytokine production and treatment response in major depressive disorder. *Neuropsychopharmacology* 2000; 22: 370–379.
173. Simen BB, Duman CH, Simen AA, Duman RS. TNF alpha signaling in depression and anxiety: behavioral consequences of individual receptor targeting. *Biol Psychiatry* 2006; 59: 775–785.
174. Tyring S, Gottlieb A, Papp K, et al. Etanercept and clinical outcomes, fatigue, and depression in psoriasis: double-blind placebo-controlled randomised phase III trial. *Lancet* 2006; 367: 29–35.
175. Nomura H, Shiosaka S, Tohyama M. Distribution of substance P-like immunoreactive structures in the brainstem of the adult human brain: an immunocytochemical study. *Brain Res* 1987; 404: 365–370.
176. Chawla MK, Gutierrez GM, Young WS, et al. Localization of neurons expressing substance P and neurokinin B gene transcripts in the human hypothalamus and basal forebrain. *J Comp Neurol* 1997; 384: 429–442.
177. Caberlotto L, Hurd YL, Murdock P, et al. Neurokinin 1 receptor and relative abundance of the short and long isoforms in the human brain. *Eur J Neurosci* 2003; 17: 1736–1746.
178. Ebner K, Rupniak NM, Saria A, Singewald N. Substance P in the medial amygdala: emotional stress-sensitive release and modulation of anxiety-related behavior in rats. *Proc Natl Acad Sci U S A* 2004; 101: 4280–4285.
179. Teixeira RM, Santos AR, Ribeiro SJ, et al. Effects of central administration of tachykinin receptor agonists and antagonists on plus-maze behavior in mice. *Eur J Pharmacol* 1996; 311: 7–14.
180. Oblin A, Zivkovic B, Bartholini G. Involvement of the D-2 dopamine receptor in the neuroleptic-induced decrease in nigral substance P. *Eur J Pharmacol* 1984; 105: 175–177.
181. Shirayama Y, Mitsushio H, Takashima M, et al. Reduction of substance P after chronic antidepressants treatment in the striatum, substantia nigra and amygdala of the rat. *Brain Res* 1996; 739: 70–78.
182. Bilkei-Gorzo A, Racz I, Michel K, Zimmer A. Diminished anxiety- and depression-related behaviors in mice with selective deletion of the Tac1 gene. *J Neurosci* 2002; 22: 10,046–10,052.
183. Morcuende S, Gadd CA, Peters M, et al. Increased neurogenesis and brain-derived neurotrophic factor in neurokinin-1 receptor gene knockout mice. *Eur J Neurosci* 2003; 18: 1828–1836.
184. Park SW, Yan YP, Satriotomo I, et al. Substance P is a promoter of adult neural progenitor cell proliferation under normal and ischemic conditions. *J Neurosurg* 2007; 107: 593–599.
185. Guiard BP, Przybylski C, Guilloux JP, et al. Blockade of substance P (neurokinin 1) receptors enhances extracellular serotonin when combined with a selective serotonin reuptake inhibitor: an in vivo microdialysis study in mice. *J Neurochem* 2004; 89: 54–63.
186. Guiard BP, Froger N, Hamon M, et al. Sustained pharmacological blockade of NK1 substance P receptors causes functional desensitization of dorsal raphe 5-HT 1A autoreceptors in mice. *J Neurochem* 2005; 95: 1713–1723.
187. Kramer MS, Cutler N, Feighner J, et al. Distinct mechanism for antidepressant activity by blockade of central substance P receptors. *Science* 1998; 281: 1640–1645.
188. Lieb K, Fiebich BL, Herpfer I, et al. No modulatory effect of neurokinin-1 receptor antagonists on serotonin uptake in human and rat brain synaptosomes. *Eur Neuropsychopharmacol* 2005; 15: 641–646.
189. Geracioti TD Jr, Carpenter LL, Owens MJ, et al. Elevated cerebrospinal fluid substance p concentrations in posttraumatic stress disorder and major depression. *Am J Psychiatry* 2006; 163: 637–643.
190. Rimon R, Le Greves P, Nyberg F, et al. Elevation of substance P-like peptides in the CSF of psychiatric patients. *Biol Psychiatry* 1984; 19: 509–516.
191. Berrettini WH, Nurnberger JI Jr, Gold PW, et al. Neuropeptides in human cerebrospinal fluid. *Life Sci* 1985; 37: 1265–1270.
192. Deuschle M, Sander P, Herpfer I, et al. Substance P in serum and cerebrospinal fluid of depressed patients: no effect of antidepressant treatment. *Psychiatry Res* 2005; 136: 1–6.
193. Bondy B, Baghai TC, Minov C, et al. Substance P serum levels are increased in major depression: preliminary results. *Biol Psychiatry* 2003; 53: 538–542.
194. Keller M, Montgomery S, Ball W, et al. Lack of efficacy of the substance p (neurokinin1 receptor) antagonist aprepitant in the treatment of major depressive disorder. *Biol Psychiatry* 2006; 59: 216–223.
195. Hirose S, Ashby CR Jr. An open pilot study combining risperidone and a selective serotonin reuptake inhibitor as initial antidepressant therapy. *J Clin Psychiatry* 2002; 63: 733–736.
196. Papakostas GI, Petersen TJ, Kinrys G, et al. Aripiprazole augmentation of selective serotonin reuptake inhibitors for treatment-resistant major

depressive disorder. *J Clin Psychiatry* 2005; 66: 1326–1330.

197. McIntyre RS, Soczynska JK, Woldeyohannes HO, et al. Aripiprazole: pharmacology and evidence in bipolar disorder. *Expert Opin Pharmacother* 2007; 8: 1001–1009.
198. Trivedi MH, Fava M, Wisniewski SR, et al. Medication augmentation after the failure of SSRIs for depression. *N Engl J Med* 2006; 354: 1243–1252.
199. Skolnick P, Popik P, Janowsky A, et al. "Broad spectrum" antidepressants: is more better for the treatment of depression? *Life Sci* 2003; 73: 3175–3179.
200. Popik P, Krawczyk M, Golembiowska K, et al. Pharmacological profile of the "triple" monoamine neurotransmitter uptake inhibitor, DOV 102,677. *Cell Mol Neurobiol* 2006; 26: 857–873.
201. Pandi-Perumal SR, Zisapel N, Srinivasan V, Cardinali DP. Melatonin and sleep in aging population. *Exp Gerontol* 2005; 40: 911–925.
202. Srinivasan V, Smits M, Spence W, et al. Melatonin in mood disorders. *World J Biol Psychiatry* 2006; 7: 138–151.
203. Wetterberg L, Bergiannaki JD, Paparrigopoulos T, et al. Normative melatonin excretion: a multinational study. *Psychoneuroendocrinology* 1999; 24: 209–226.
204. Rubin RT, Heist EK, McGeoy SS, et al. Neuroendocrine aspects of primary endogenous depression. XI. Serum melatonin measures in patients and matched control subjects. *Arch Gen Psychiatry* 1992; 49: 558–567.
205. Thompson C, Mezey G, Corn T, et al. The effect of desipramine upon melatonin and cortisol secretion in depressed and normal subjects. *Br J Psychiatry* 1985; 147: 389–393.
206. Carman JS, Post RM, Buswell R, Goodwin FK. Negative effects of melatonin on depression. *Am J Psychiatry* 1976; 133: 1181–1186.
207. Dolberg OT, Hirschmann S, Grunhaus L. Melatonin for the treatment of sleep disturbances in major depressive disorder. *Am J Psychiatry* 1998; 155: 1119–1121.
208. Banasr M, Soumier A, Hery M, et al. Agomelatine, a new antidepressant, induces regional changes in hippocampal neurogenesis. *Biol Psychiatry* 2006; 59: 1087–1096.
209. Millan MJ, Brocco M, Gobert A, Dekeyne A. Anxiolytic properties of agomelatine, an antidepressant with melatoninergic and serotonergic properties: role of 5-HT2C receptor blockade. *Psychopharmacology (Berl)* 2005; 177: 448–458.
210. Papp M, Gruca P, Boyer PA, Mocaer E. Effect of agomelatine in the chronic mild stress model of depression in the rat. *Neuropsychopharmacology* 2003; 28: 694–703.
211. Loo H, Hale A, D'Haenen H. Determination of the dose of agomelatine, a melatoninergic agonist and selective 5-HT(2C) antagonist, in the treatment of major depressive disorder: a placebo-controlled dose range study. *Int Clin Psychopharmacol* 2002; 17: 239–247.
212. Montgomery SA, Kasper S. Severe depression and antidepressants: focus on a pooled analysis of placebo-controlled studies on agomelatine. *Int Clin Psychopharmacol* 2007; 22: 283–291.
213. Kennedy SH, Emsley R. Placebo-controlled trial of agomelatine in the treatment of major depressive disorder. *Eur Neuropsychopharmacol* 2006; 16: 93–100.
214. Zupancic M, Guilleminault C. Agomelatine: a preliminary review of a new antidepressant. *CNS Drugs* 2006; 20: 981–992.
215. Tanskanen A, Hibbeln JR, Hintikka J, et al. Fish consumption, depression, and suicidality in a general population. *Arch Gen Psychiatry* 2001; 58: 512–513.
216. Adams PB, Lawson S, Sanigorski A, Sinclair AJ. Arachidonic acid to eicosapentaenoic acid ratio in blood correlates positively with clinical symptoms of depression. *Lipids* 1996; 31(suppl): S157–S161.
217. Lin PY, Su KP. A meta-analytic review of double-blind, placebo-controlled trials of antidepressant efficacy of omega-3 fatty acids. *J Clin Psychiatry* 2007; 68: 1056–1061.
218. Janowsky DS, el-Yousef MK, Davis JM, Sekerke HJ. A cholinergic-adrenergic hypothesis of mania and depression. *Lancet* 1972; 2: 632–635.
219. Picciotto MR, Brunzell DH, Caldarone BJ. Effect of nicotine and nicotinic receptors on anxiety and depression. *Neuroreport* 2002; 13: 1097–1106.
220. Janowsky DS, Overstreet DH. Cholinergic dysfunction in depression. *Pharmacol Toxicol* 1990; 66(suppl 3): 100–111.
221. Shytle RD, Silver AA, Lukas RJ, et al. Nicotinic acetylcholine receptors as targets for antidepressants. *Mol Psychiatry* 2002; 7: 525–535.
222. Laje RP, Berman JA, Glassman AH. Depression and nicotine: preclinical and clinical evidence for common mechanisms. *Curr Psychiatry Rep* 2001; 3: 470–474.
223. Killen JD, Fortmann SP, Schatzberg A, et al. Onset of major depression during treatment for nicotine dependence. *Addict Behav* 2003; 28: 461–470.
224. Rabenstein RL, Caldarone BJ, Picciotto MR. The nicotinic antagonist mecamylamine has antidepressant-like effects in wild-type but not beta2- or alpha7-nicotinic acetylcholine receptor subunit knockout mice. *Psychopharmacology (Berl)* 2006; 189: 395–401.
225. Mineur YS, Somenzi O, Picciotto MR. Cytisine, a partial agonist of high-affinity nicotinic acetylcholine receptors, has antidepressant-like properties in male C57BL/6J mice. *Neuropharmacology* 2007; 52: 1256–1262.
226. Furey ML, Drevets WC. Antidepressant efficacy of the antimuscarinic drug scopolamine: a randomized, placebo-controlled clinical trial. *Arch Gen Psychiatry* 2006; 63: 1121–1129.
227. Wisniewski SR, Fava M, Trivedi MH, et al. Acceptability of second-step treatments to depressed outpatients: a STAR*D report. *Am J Psychiatry* 2007; 164: 753–760.
228. Taylor MJ, Freemantle N, Geddes JR, Bhagwagar Z. Early onset of selective serotonin reuptake inhibitor antidepressant action: systematic review and meta-analysis. *Arch Gen Psychiatry* 2006; 63: 1217–1223.
229. Bradley RG, Binder EB, Epstein MP, et al. Influence of child abuse on adult depression: moderation by the corticotropin-releasing hormone receptor gene. *Arch Gen Psychiatry*. 2008; Feb;65(2): 190–200.
230. Takahashi LK, Ho SP, Livanov V, et al. Antagonism of CRF(2) receptors produces anxiolytic behavior in animal models of anxiety. *Brain Res*. 2001; Jun 1;902(2): 135–42.

Bipolar Disorder: Candidate Drug Targets

Carlos A. Zarate Jr and Husseini K. Manji

Laboratory of Molecular Pathophysiology and Experimental Therapeutics,
Mood and Anxiety Disorders Research Program, National Institute of Mental Health, Bethesda, MD

ABSTRACT

Current pharmacotherapy for bipolar disorder is generally unsatisfactory for a large number of patients. Even with adequate modern bipolar pharmacological therapies, many afflicted individuals continue to have persistent mood episode relapses, residual symptoms, functional impairment, and psychosocial disability. Creating novel therapeutics for bipolar disorder is urgently needed. Promising drug targets and compounds for bipolar disorder worthy of further study include both systems and intracellular pathways and targets. Specifically, the purinergic system, the dynorphin opioid neuropeptide system, the cholinergic system (muscarinic and nicotinic systems), the melatonin and serotonin [5-hydroxytryptamine receptor 2C] system, the glutamatergic system, and the hypothalamic-pituitary adrenal axis have all been implicated. Intracellular pathways and targets worthy of further study include glycogen synthase kinase-3 protein, protein kinase C, and the arachidonic acid cascade. ***Mt Sinai J Med 75:226–247, 2008.*** © *2008 Mount Sinai School of Medicine*

Key Words: antidepressant, bipolar disorder, depression, glutamate, mania, plasticity, treatment.

Bipolar disorder (BPD) is one of the most severe illnesses of the major mental disorders and ranks in the top 10 causes of medical disability. It is very common, having a lifetime prevalence of approximately 4.4% (BPD I, 1.0%, and BP II, 1.1%) in the United States.[1] BPD is a complex illness encompassing varying degrees of fluctuating disturbances of emotions, behavior, thought, cognition, and hedonic and motoric drive over the course of a person's life span. Such varied clinical syndromes are usually encapsulated into episodes for diagnostic and treatment purposes (manic, mixed, hypomanic, and depressive episodes). As a result, the development of BPD therapies occurs first for the acute phases of the illness (manic, mixed, and depressive episodes) and then for the maintenance phase of treatment, even though controlling relapses is the most important aspect in the treatment of BPD. Pharmacologically, the greatest success has been in the treatment of acute manic episodes. There are now many antimanic agents available for clinical use, although a sizable proportion of patients fail to respond to or tolerate these treatments.[2] However, for acute depressive episodes and for maintenance treatment, few treatments have been proven to be effective; this is especially of concern because depressive episodes and symptoms and depressive relapse dominate the longitudinal course of BPD.[3]

Recent large-scale National Institute of Mental Health–funded studies have explored the effectiveness of our standard treatments for patients with recurrent mood disorders, and the news is of concern. The Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study found that less than one-third of patients with major depressive disorder achieved remission with an adequate trial of a standard antidepressant after approximately 10 to 14 weeks of treatment. For BPD depression, the use of adjunctive, standard antidepressant medication, in comparison with the use of mood stabilizers, was not associated with increased efficacy after 26 weeks.[4–6] Except for lithium, all available Food and Drug Administration (FDA)–approved treatments for BPD fall into the category of anticonvulsant or antipsychotic drugs and were originally developed to treat other conditions.[7] Until recently, no drug has been developed specifically for the treatment of BPD on the basis of an understanding of the neurobiological basis of the illness or of the mechanism of action of existing effective medications. The lack of novel treatments for BPD is undoubtedly due in part to the

Address Correspondence to:

Carlos A. Zarate
Mark O. Hatfield Clinical
Research Center
10 Center Drive
Unit 7 SE, Room 7-3445
Bethesda, MD 20892
Email: zaratec@mail.nih.gov

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20042

© 2008 Mount Sinai School of Medicine. †This article is a US government work and, as such, is in the public domain in the

complexity of studying this illness (eg, the difficulty in recruiting patients, natural course of the disease, placebo effects, and high rates of dropout).

In order to determine what neurotransmitter systems/intracellular pathways might be relevant for developing drugs for BPD, we conducted a Medline and web-based search (1990 to October 2007), looking for examples of compounds or drugs that met 1 or more of the following criteria: (1) antimanic effects in humans or beneficial effects on irritability, hyperactivity, or mood; (2) antidepressant properties in bipolar depression; and (3) antidepressant-like properties in animal models and either "antimanic-like" properties in animal models[8] or antipsychotic-like properties in either humans or animal models of psychosis (eg, prepulse inhibition). Antipsychotic or anticonvulsant drugs were excluded, except for those with predominant effects on α-amino-3-hydroxy-5-methyl-4-isoxazolepropionate (AMPA) receptors. A review of these agents is not included here because this topic has been discussed extensively elsewhere and because it is likely that useful drugs from these classes of compounds will continue to become available for the treatment of BPD but are likely to be comparable in efficacy to existing medications. Finally, it should be noted that the extrapolation of findings from animal studies to humans in the absence of reliable and valid animal models of BPD must be interpreted with caution.

Here we review the drug targets and compounds for BPD meeting these criteria. Several systems are worthy of further study, including (1) the purinergic system, (2) the dynorphin opioid neuropeptide system, (3) the cholinergic system (muscarinic and nicotinic systems), (4) the melatonin and serotonin [5-hydroxytryptamine receptor 2C (5-HT2C)] system, (5) the glutamatergic system, and (6) the hypothalamic-pituitary adrenal (HPA) axis. In addition, several intracellular pathways and targets merit further attention, including (1) glycogen synthase kinase-3 (GSK-3) protein, (2) protein kinase C (PKC), (3) the arachidonic acid (AA) cascade, and (4) other candidates. It is important to note that most of the drugs reviewed here are proof-of-concept studies, some with very small sample sizes. Thus, the generalizability of such preliminary findings to current clinical practice patterns would be premature.

SYSTEMS WORTHY OF FURTHER STUDY IN BPD

Purinergic System

Purines play an essential role in energy metabolism and are regulators of neurotransmission (adenosine triphosphate and adenosine); adenosine is a widespread neuromodulator acting mostly through adenosine 1 and 2A receptors. Uric acid is the ultimate step in the metabolism of the purinergic system. In the 19th century, uric acid "diathesis," a predisposition to the accumulation of urea in the body, was believed to cause rheumatism, cardiac disease, and mental illness.[9] Because lithium urate was found to dissolve urate stones, it was believed that it could be helpful in the treatment of these conditions. In 1949, Cade[10] injected lithium urate into guinea pigs, noted that it had a calming effect, and reasoned that it would be helpful in calming patients with mania. Anumonye *et al.*[11] reported that remission in mania was associated with the increased excretion of uric acid. Subsequently, it was hypothesized that a purinergic dysfunction might be involved in the neurobiology of mania,[12] and genetic data implicate purinergic dysfunction in BPD.[13,14]

The avoidance of adenosine antagonists such as caffeine has been recommended for patients with BPD because of its potential to cause irritability and disrupt the sleep-wake cycle; the latter is a common reason for manic relapse. A case of secondary mania caused by caffeine has been reported.[15] Adenosine agonists have been reported to have sedative, anticonvulsant, anti-aggressive, and antipsychotic properties in animals.[16] With respect to other purinergic modulators, allopurinol has been used for many years for the treatment of gout; it acts by inhibiting xanthine oxidase, a key step in the production of uric acid.[17] Case reports suggest that allopurinol might be effective in the treatment of mania and hyperuricemia.[18] Recently, 2 large, placebo-controlled trials confirmed that the addition of allopurinol to ongoing antimanic/mood-stabilizer therapies resulted in significant antimanic effects. In the first study,[19] 82 subjects were randomized to either blinded allopurinol (300 mg/day) or placebo added to lithium plus haloperidol for 8 weeks. Post-hoc comparisons showed significant improvement as early as day 7 on the Young Mania Rating Scale (YMRS), and the difference between the 2 groups was also significant at the endpoint (8 weeks). Side effects for the 2 groups were comparable. The second study was a 4-week, double-blind, placebo-controlled study involving 150 subjects with acute bipolar mania. The study compared allopurinol (600 mg/day), dipyridamole (200 mg day), and placebo added to lithium.[20] Further large controlled studies with more selective modulators of the purinergic system are needed to determine what aspects of the purinergic system are relevant to antimanic effects.

Dynorphin Opioid Neuropeptide System

The dynorphin opioid neuropeptide system is involved in mood, motor, cognitive, and endocrine functions. A number of preclinical studies support the evidence of the opioid system's putative involvement in depression. There are 3 well-defined types of opioid receptors: μ, κ, and δ. All these types of opioid receptors have been implicated to different degrees in major depression. A significant reduction (37%–38%) of the prodynorphin messenger RNA (mRNA) expression levels in the amygdalohippocampal area and in the parvicellular division of the accessory basal area in patients with BPD was found.[21]

Kappa Opioid Receptors Kappa opiate agonists such as U50,488 produce analgesia and diuresis and show antipruritic activity.[22] Selective kappa opiate antagonists are being explored for their effects in the treatment of a wide variety of conditions, including cocaine addiction[23] and feeding behavior abnormalities,[24] and have been proposed as a treatment for psychosis.[25] The activation of kappa opiate receptors has depressogenic effects in both animals[26] and humans.[27] The blockade of kappa opiate receptors results in antidepressant-like properties in animals.[28] Recently, evidence of the antidepressant-like properties of the kappa opioid antagonist MCL-144B in the forced swim test (FST) has been reported.[29] Thus, it stands to reason that kappa opiate agonists could have antimanic effects. However, it has long been recognized that centrally acting kappa opiate receptor agonists might have limited usefulness in humans because of the psychotomimetic and dysphoric actions that can occur with their use.[30,31]

Pentazocine No selective kappa agonists are available for testing in humans. However, the analgesic pentazocine (Talwin) penetrates the blood-brain barrier and is a partial agonist at the kappa opiate receptor. In a recent open-label study involving 10 inpatients in the manic phase of BPD with a YMRS score of ≥14, adjunctive pentazocine significantly improved manic symptoms without inducing depression.[32] In this inpatient study, subjects received two 50-mg doses of pentazocine 2 hours apart. Symptoms of mania were reported to be reduced 1 hour after each dose, >44% after the first dose and 41% after the second dose, and this was not due to sedation. Overall, the study medication was well tolerated; no subject complained of dysphoria when self-rating mood, and significant adverse events or psychotomimetic effects were not observed. It is important to emphasize that this was a small, uncontrolled study. At this time, pentazocine is not advocated for clinical use because excessive doses could lead to serious intoxication[33] and because there have been reports that it induces symptoms of depression and "gloominess" in healthy volunteers.[34] However, it will be interesting to follow up this preliminary study with controlled, proof-of-principle studies using more selective kappa opiate agonists in the treatment of mania in order to determine the relevance of kappa opioid receptors in BPD.

Salvinorin Salvinorin-A is a recreational drug derived from the *Salvia divinorum* plant, a member of the sage family.[35] It is a naturally occurring hallucinogen identified to be a highly selective full kappa opioid receptor agonist.[25] A recent study[26] found that salvinorin-A induced depressive-like behaviors in the FST (increased immobility) and intracranial-stimulation test. In this report, the investigators noted that, at the dose of salvinorin-A that caused depressive-like effects, extracellular concentrations of dopamine, but not serotonin, within the nucleus accumbens were reduced. A greater understanding of the mode of action of salvinorin-A will undoubtedly lead to a better understanding of the kappaergic system in human brain function and to the development of a series of novel compounds for potential therapeutic use in neuropsychiatric conditions.[36] It should also be noted that the potential for abuse and/or withdrawal of these types of compounds will need to be further studied before they can be of clinical utility.

Cholinergic System

The cholinergic system has long been studied as an important aspect of the pathophysiology of depression. Several decades ago, Janowsky *et al.*[37] introduced the "cholinergic-adrenergic hypothesis" of depression and mania. They postulated that dysfunctions of the cholinergic-adrenergic balance might be associated with the pathophysiology of mood disorders. Although no evidence for the association between 19 cholinergic genes and BPD was found,[38] other evidence supports the role of this system in mood disorders. For instance, rats bred selectively for increased sensitivity of muscarinic receptors demonstrated behaviors that are similar to those seen in patients with depression, such as lethargy, anhedonia, and behavioral despair.[39] In humans, enhanced cholinergic activity induced a worsening of symptoms in patients with unipolar depression.[37] Furthermore, neuroendocrine and pupillary responses to cholinergic activity are augmented in depressed subjects[40] and decreased in manic subjects.[41] Improvement in mania with lithium and valproate is linked with normalization of pupillary responses.[41]

In addition, a recent positron emission tomography imaging study reported reduced muscarinic type 2 receptor binding in the anterior cingulate cortex in subjects with BPD.[42] Further supporting the importance of the role of the cholinergic system in BPD is the fact that small, controlled trials of physostigmine, a short-acting cholinesterase inhibitor, led to rapid but only temporary decreases in symptoms of mania after single or multiple intravenous injections.[43,44] A more recent study found that the long-acting cholinesterase inhibitor donepezil had rapid antimanic effects. In this open case series, donepezil (5–10 mg/day) was added to ongoing mood-stabilizer treatment and led to significant improvement in 6 of 11 patients with treatment-resistant mania.[45] However, a recent 6-week, double-blind, placebo-controlled trial failed to find that adjunctive donepezil (5–10 mg/day) was effective in the treatment of refractory manic symptoms.[46] It is possible that the small size of the sample (12 subjects randomized) could have led to a false-negative result, although YMRS scores at the endpoint were significantly higher in the donepezil group than the placebo group.

Until recently, only a handful of uncontrolled studies have suggested that anticholinergic drugs might have antidepressant efficacy. The antidepressant effects of tricyclic antidepressants were believed to be largely due to their anticholinergic properties. Earlier work by Kasper and colleagues[47] described the antidepressant properties of the anticholinergic drug biperiden in 10 severely depressed inpatients. More recently, Furey and Drevets[48] reported on 2 studies that unexpectedly found that the antimuscarinic drug scopolamine had antidepressant properties in subjects with unipolar and bipolar depression; these effects were rapid, occurring within 3 to 5 days. In the first of the 2 studies, 4 testing sessions were performed in random order under double-blind conditions, during which subjects received a 15-minute intravenous infusion of a saline placebo and 3 doses of scopolamine hydrobromide (2.0, 3.0, and 4.0 μg/kg). The second study involved 7 sessions in which subjects received a 15-minute intravenous infusion of a placebo saline solution or scopolamine hydrobromide (4.0 μg/kg). The number of subjects with BPD was not specified in the first study, but there were 9 subjects with BPD in the second study. Analysis indicates that patients with BPD and major depressive disorder separately showed significant reductions in Montgomery-Asberg Depression Rating Scale (MADRS) scores when the endpoint was compared with the baseline. The authors reported that the antidepressant effects seen in the study with scopolamine were not simply due to euphoria; previous reports had suggested that anticholinergic drugs had this property.[49] In this study, only 1 patient developed euphoria on scopolamine. Furthermore, no increase in YMRS scores occurred in patients with BPD during scopolamine treatment when the baseline was compared with the study endpoint. In animals, scopolamine was reported to produce a significant dose-dependent decrease in prepulse inhibition, but it had no effect on startle amplitude.[50]

Because anticholinergic side effects are a common complaint and reason for discontinuing tricyclic antidepressants in depressed patients, novel compounds currently in development that target the cholinergic system would need to factor in this potential problem. Further controlled short- and long-term studies are warranted to determine the efficacy, safety, and tolerability of anticholinergic compounds in mood disorders.

Nicotinic Acetylcholine Receptor (nAChR) System

The nAChR system is one of the best characterized neurotransmitter systems. To date, 12 neuronal nAChR subunits have been identified (α2–10 and β2–4).[51] The different pentameric nAChR subtypes can result from these subunit combinations and are distinguished on the basis of their affinities for nicotine (low or high) and other nAChR ligands. The nAChRs expressed in the mammalian brain with low affinity for the nicotine receptor appear to be α7 homomers, whereas those with a high affinity for nicotine have primarily α4 and β2 subunits.[52] A link between nicotine and depression has been suggested by epidemiological studies of smokers.[53] In animals, nicotine has been reported to have mood-elevating properties (hedonic), and its withdrawal induces an anhedonic state.[54] Indirect evidence for the involvement of nAChRs in depression was provided by the Flinders sensitive line, a genetic animal model of depression in rats that could be reversed by nicotine treatment.[55] Similarly, nicotine has been found to have antidepressant-like properties in a learned helplessness test in rats.[56] Cytisine, a partial agonist of high-affinity nAChRs (a partial agonist of $\alpha4/\beta2$ and a full agonist at $\alpha3/\beta4$) was recently found to have antidepressant-like properties in male C57BL/6J mice.[57]

Evidence for the involvement of the nicotinic system in BPD includes the fact that mRNA levels of $\alpha7$ (located at chromosome 15a13-14) and $\alpha7$-like genes from the postmortem prefrontal cortex were significantly expressed in the postmortem tissue of subjects with BPD, but not in postmortem tissue of patients with schizophrenia or healthy controls.[58]

In a recent 4-week, double-blind study, 11 nonsmokers with depressive symptoms (Center for Epidemiological Studies Depression Scale ≥ 10) were randomized to either transdermal nicotine (3.5–7.0 mg/day) or placebo. Nicotine induced significant antidepressant effects in comparison with placebo at day 8 but not on days 21 and 28. Reported side effects were infrequent and minimal.[59] These data appear to indicate that nicotine could play a role in major depression; however, the use of chronic nicotine as a standard treatment for depression is limited by obvious health risks and its side-effect profile (eg, nausea and sympathomimetic actions). The unwanted side effects are probably mediated through peripheral nicotinic receptors, and for that reason, studies with nicotinic subtype inhibitors are underway.

SIB-1508Y is a selective $\alpha4/\beta2$ nAChR agonist and has been found to have antidepressant-like properties in the learned helplessness model of depression in rats.[60] As with fluoxetine and imipramine, a dose-dependent study of subchronic treatment with SIB-1508Y reversed the escape deficit in the learned helplessness model, and this effect was still apparent 1 week later.

In terms of nicotinic antagonists, a previous study found that comorbid BPD was improved in 2 Tourette's syndrome patients treated with mecamylamine (2.5–7.5 mg/day). Mecamylamine blocks nicotinic receptors and was noted to have mood-stabilizing properties in these 2 patients; manic symptoms became apparent only upon cessation of mecamylamine treatment.[62] Furthermore, a more recent study also found mecamylamine to have antidepressant and mood-stabilizing properties.[63] In this 8-week, double-blind, placebo-controlled trial, mecamylamine (2.5 to 7.5 mg/day) significantly decreased sudden mood changes in children and adolescents (8–17 years old) with Tourette's disorder and other comorbid disorders (attention deficit hyperactivity disorder, obsessive-compulsive disorder, and hypomania). In a completer's analysis, significant differences that favored mecamylamine over placebo in the comorbid major depression group included "irritable," "sudden mood changes," "inattention," "restless" or hyper," "tense, anxious, nervous," and "impulsive." In the hypomania comorbid group, significant differences that favored mecamylamine over placebo included "restless or hyper" and "depressed or uninterested in most things." Although there are limitations with the aforementioned reports (including small sample size, overlapping comorbid diagnoses, and not direct testing in patients with BPD), the preliminary nature of the data indicating beneficial improvement of symptoms commonly seen in BPD with mecamylamine suggests that controlled trials with selective inhibitors of nAChRs in patients with mood disorders should be considered. These results complement the preclinical findings of "antimanic-like properties," in that mecamylamine attenuates both ephedrine-induced[64] and quinpirole-induced hyperactivity in rats.[65]

In addition, mecamylamine was found in a dose-response study to significantly decrease immobility time the FST and tail suspension test at the dose of 1.0 mg/kg without altering baseline locomotor activity. These effects appear to be dependent on both $\beta2$ and $\alpha7$ subunits of the NAChR, as mice lacking these subunits failed to show evidence of the antidepressant-like properties61. Similarly to the opiate agonists, the potential for abuse and/or withdrawal symptoms from these types of compounds needs further study.

Melatonin and Serotonin (5-HT2C Receptor) System

The pineal hormone melatonin produces most of its biological effects via G protein–coupled melatonin receptors (MT1 and MT2). In mammalian tissues, these receptors are particularly expressed in the brain. The cyclical nature of BPD, the varied fluctuations in its symptomatology, and the existence of disturbed sleep-wake rhythms all suggest that dysfunction of the circadian system may underlie the pathophysiology of this disorder. Supersensitivity of the melatonin-suppressing effects of light has also been reported in patients with BPD, in nonaffected offspring of probands with BPD, and in monozygotic twins discordant for BPD.[66–68] However, Nurnberger and colleagues[69] did not confirm that melatonin suppression by light occurred in euthymic bipolar patients. Thompson and colleagues[70] found a significant association of the δ 502–505 polymorphism in GPR50 (also known as H9, melatonin-related receptor, or ML1X and located on Xq28) and susceptibility to BPD in a population from the Southeast Scotland. However, this finding was not replicated in a Northern Swedish association sample.[71]

Interestingly, low doses of both lithium and valproate have been reported to reduce melatonin light sensitivity but not its overall synthesis in healthy volunteers.[72,73] There are no controlled studies with melatonin in patients with BPD. Case reports and series indicate mixed results.[74,75]

More recently, the availability of agomelatine, a potent agonist of melatonin MT1 and MT2 receptors, has allowed researchers to test the relevance of melatonin system treatments for mood disorders. However, agomelatine is not selective for MT1 and MT2, as it is also a 5-HT2C antagonist; the latter property has also been implicated in the mechanism of action of atypical antipsychotic drugs with thymoleptic properties.[76] In vivo studies indicate that agomelatine increases both norepinephrine and dopamine in the frontal cortex. In addition to increasing these neurotransmitters, chronic agomelatine treatment (3 weeks) resulted in increased cell proliferation and neurogenesis in the ventral dentate gyrus as well as increased survival of these newly formed cells.[77] Furthermore, similar to lithium,[78] agomelatine is able to resynchronize a disrupted circadian rhythm and has circadian phase-advancement properties.[79,80] Agomelatine is effective in animal models of depression (specifically, in the FST, chronic mild stress test, and learned helplessness model)[81,82] and anxiety (the social interaction test and Vogel conflict test).[83]

In 3 large, multicenter, multinational, placebo-controlled, short-term studies in major depression, agomelatine was found to be a clinically effective and well-tolerated antidepressant.[84–86] In the study by Loo and colleagues,[84] some patients met criteria for bipolar II disorder (depressed); an analysis was not presented separately for patients with this type of diagnosis. Agomelatine appears to improve sleep quality and ease of falling asleep, as measured subjectively in depressed patients.[87] Agomelatine also shows a low liability for hypomania or mania in the several studies conducted to date,[88] which might explain the impetus to study it in BPD.

In a recent study, 21 patients with bipolar I disorder who were experiencing a major depressive episode and had a 17-Item Hamilton Depression Scale score of ≥ 18 were studied. All patients received agomelatine (25 mg/day) for 6 weeks with a possible extension of up to 46 weeks in combination with either lithium (n = 14) or valpromide (n = 7). According to intent-to-treat data, 81% of patients met criteria for marked improvement at the study endpoint, and 47% responded as early as the first week of treatment. Afterwards, 19 patients entered the 1-year extension phase of the study, and 11 completed it. There were no dropouts due to adverse events during the acute phase of treatment (6 weeks), although 6 patients experienced serious adverse events during the 1-year period. Three lithium-treated patients experienced manic or hypomanic episodes during the optional extension period, one of which was treatment-related.[89] One interesting characteristic of the drug is that apparently its abrupt cessation does not result in discontinuation symptoms.[90]

Glutamatergic System

Increasingly, the glutamatergic system is being recognized as a likely contributor to impairments in brain neuroplasticity and cellular resilience observed in patients with BPD. The preclinical evidence supporting the role of glutamate in the pathophysiology of depression or mechanism of action has been summarized elsewhere.[91–93] Glutamate is the major excitatory synaptic neurotransmitter in the brain. Its crucial functions include mediating neurotransmission across excitatory synapses and modulating various physiological functions in the mammalian central nervous system such as synaptic plasticity, learning, and memory.[94–97] Excessive concentrations of glutamate are hypothesized to be involved in the etiopathophysiology of several neurodegenerative illnesses. Consequently, several drugs have been created in an attempt to modulate these abnormal concentrations. Evidence that these types of compounds are neuroprotective in humans with ischemic/traumatic or neurodegenerative disease is still pending.

Several of the glutamatergic compounds being studied as neuroprotective agents either have been or are now undergoing testing in "proof-of-concept" studies in patients with severe mood disorders.[91,98] The glutamatergic modulators that are being developed target either the glutamate receptors [*N*-methyl-D-aspartate (NMDA), AMPA, and metabotropic] directly or glutamate before it is released into the extracellular space.

Emerging data indicate that glutamate has an important role in both acute and long-term processes involved in the mode of action of antidepressants and/or mood stabilizers. Synaptic potentiation by enhancing AMPA throughput is thought to be involved in acute antidepressant response, whereas the positive neurotrophic changes resulting from glutamatergic modulators are perhaps more relevant to reducing the recurrence of mood episodes and minimizing the deleterious effects of chronic aberrant neurobiology.

Glutamate Release and AMPA Trafficking Riluzole (2-amino-6-trifluoromethoxy benzothiazole) is a blood-brain-penetrant glutamatergic modulator with neuroprotective and anticonvulsant properties. Riluzole is the only drug approved by the FDA for the treatment of the degenerative motor-neuron disease amyotrophic lateral sclerosis. There is mounting evidence that it has multiple effects on the glutamatergic system, including inhibition of glutamate release, enhancement of AMPA trafficking by increasing membrane insertion of AMPA subunits glutamate receptor type 1 (GluR1) and glutamate receptor type 2 (GluR2),[99] and glutamate reuptake.[91,99,100] Riluzole is also known to stimulate the synthesis of growth factors, including brain-derived neurotrophic factor (BDNF) in cultured mouse astrocytes,[101] and was recently shown to have antidepressant-like properties in animal models (G. Sanacora, unpublished data, 2008).

Several clinical studies have been conducted with riluzole. It was found to have antidepressant effects in patients with unipolar depression[102,103] as well as bipolar depression.[104] Overall, riluzole was well tolerated in these trials. These preliminary results need to be confirmed in controlled studies. In mice, pretreatment with 10 mg/kg riluzole, but not 3 mg/kg riluzole, moderately reduced amphetamine-induced hyperlocomotion but not dizocilpine-induced hyperlocomotion.[105] This characteristic suggests that it might have "antimanic-like" properties, but more studies are needed to confirm this.

Ionotropic Glutamate Receptors Three subgroups of glutamatergic ion channels have been identified on the basis of their pharmacological ability to bind different synthetic ligands: NMDA, AMPA, and kainate receptors.

NMDA Receptor Complex Data from preclinical and clinical investigations support the concept that the NMDA receptor complex may play a major role in the pathophysiology of mood disorders and the mechanism of action of antidepressants and possibly mood stabilizers (reviewed by Zarate *et al.*[91]). NMDA receptor antagonists (eg, dizocilpine and 2-amino-7-phosphonoheptanoic acid) as well as an AMPA receptor potentiator have been shown to have antidepressant properties in several animal models of depression and to induce neurogenesis in the brains of rats (reviewed by Zarate *et al.*[91] and Cameron *et al.*[106]). The next section summarizes studies investigating a partial glycine receptor agonist and NMDA antagonists in patients with mood disorders.

D-Cycloserine D-Cycloserine, an antibiotic used in the treatment of tuberculosis, is a partial agonist of the glycine recognition site of the NMDA receptor. In terms of evidence for its "antimanic properties," only preclinical data have been published. D-Cycloserine was found to inhibit the hypermobility induced by methamphetamine but not that induced by apomorphine[107] and to decrease aggressiveness in the resident-intruder test.[108] In humans, D-cycloserine is currently being tested in bipolar depression in a trial supported by the Theodore and Vada Stanley Foundation. A recent study found it to be ineffective in treatment-resistant depression.[109] However, in addition to the limitations of this study as acknowledged by the authors (eg, unevenness in treatment-resistance criteria and concomitant medications), it is notable that in animal studies, a single dose of D-cycloserine elicited a dose-dependent reduction in immobility in the Porsolt swim test, whereas multiple doses did not.[110] The investigators surmised that chronic administration of NMDA glycine partial agonists produced a behavioral tolerance putatively through adaptation of the NMDA receptor complex.

Memantine Memantine is a noncompetitive NMDA antagonist with both anticonvulsant and neuroprotective properties, and it is FDA-approved for the treatment of Alzheimer's disease. Memantine is "use-dependent," in that it blocks the NMDA receptor–associated ion channel only when the channel is open for long periods, as occurs in states of excitotoxicity. Consequently, the problematic effects of pathological concentrations of glutamate are prevented to a greater extent than the effects of physiological concentrations, which are relatively spared with memantine.[111,112] Memantine is a fairly selective NMDA receptor antagonist at doses of 5 to 20 mg/day with negligible affinity for other receptors that have been implicated in antidepressant action.

Memantine has antidepressant-like effects when used alone[113] or synergistically with imipramine[114] in the FST in rats. Although memantine showed antidepressant-like effects in preclinical studies, in humans at the doses tested (up to 20 mg/day), it was found to be devoid of significant antidepressant or antianxiety effects in a double-blind, placebo-controlled trial in patients with major depression.[115] Although no significant antidepressant effects were found within this study, it does not disprove the possibility that higher doses of memantine, augmentation strategies, or its use in different populations (eg, in BPD) may result in positive effects.

A recent open-label study involving 8 patients with major depressive disorder suggests that higher doses of memantine might be beneficial for some.[116] In addition, a recent case series involving 2 patients indicated that memantine at doses of 10 to 20 mg/day improved depressive symptoms and cognitive performance in patients with BPD when added to existing mood-stabilizer therapy.[117] Another small, open-label trial found that memantine had beneficial effects in hyperactivity and irritability in children with pervasive development disorders; these symptoms are often seen in patients with BPD.[118] However, although this finding implies that memantine could be effective in treating the core symptoms of mania, it does not offer additional support for considering its further study in BPD. Another controlled study is investigating its use as an augmentation therapy in patients with bipolar depression who have an incomplete response to lamotrigine.

In terms of "antipsychotic-like" properties in animal models, memantine has been shown to disrupt prepulse inhibition of acoustic startle in rats.[119] However, caution should be used with this information as NMDA antagonists, even those of low affinity, may have a propensity to induce psychosis, especially if used at high doses or in individuals prone to develop psychosis. Furthermore, memantine at high doses has been reported to induce seizures in kindled rats.[120]

Ketamine In contrast to the evidence suggesting that memantine does not possess antidepressant effects in patients with major depression, there is increasing proof that the higher affinity NMDA receptor antagonist, ketamine, has antidepressant effects. Two controlled studies found that ketamine resulted in rapid antidepressant effects in patients with treatment-resistant (unipolar) depression.[98,121] The effects noted in the latter study were rapid (within 2 hours), robust, and relatively sustained (lasting approximately 1 week). Because of the inherent propensity of the compound to produce cognitive deficits and psychotomimetic effects, its use at this time remains limited to the research setting. Studies with more selective subtype NMDA antagonists are underway in order to determine whether these have antidepressant effects that can occur safely without causing ketamine's undesirable side effects.

AMPA Receptors AMPA receptors are ionotropic receptors implicated in learning and memory that mediate the fast component of excitatory neurotransmission. Numerous classes of compounds modulate AMPA receptors by binding to their allosteric sites and are termed AMPA receptor positive modulators or AMPA receptor potentiators (ARPs). ARPs regulate the AMPA receptors indirectly by slowing the receptor desensitization rate and/or deactivation in the presence of an agonist (eg, AMPA and glutamate; see Borges and Dingledine[122] and Bleakman and Lodge[123] for reviews). Positive modulators of these receptors, the AMPAkines, allosterically produce positive modulation of these receptors. These compounds are under active investigation as treatments for cognition, depression, anxiety, stroke, and Parkinson's disease (reviewed by Black[124]). Chronic treatment with traditional antidepressants increases the expression of AMPA receptors in hippocampal membranes[125] and the phosphorylation of AMPA receptor subunits.[126]

LY392098, an ARP, was found to have "antidepressant-like" properties in the FST and in the tail suspension test. LY392098 alone dose-dependently reduced immobility in a manner similar to that of classic antidepressants, and at subthreshold doses, it potentiated the antidepressant effects of conventional antidepressants.[127] In contrast to conventional antidepressants, this group of compounds does not affect the extracellular concentration of monoamines.[128] In primary neuronal cultures, LY392098 increased BDNF mRNA,[129] which has been implicated in the mechanism of action of many currently available antidepressants.[130,131] Another ARP that is being investigated in depression is S18986, which has also been reported to increase BDNF expression.[129,132]

In terms of the effects of standard treatments for BPD (ie, lithium, valproate, and lamotrigine) on AMPA receptors, those agents with a predominantly antidepressant profile, namely, lamotrigine and riluzole, significantly enhanced the surface expression of GluR1 and GluR2 in a time- and dose-dependent manner in cultured hippocampal neurons. In contrast, the predominantly antimanic agents lithium and valproate significantly reduced surface expression of GluR1 and GluR2.[99,133] These findings imply that regulation of GluR1/2 surface levels and function may be involved in the different clinical profile of anticonvulsants and suggest that drugs that mimic these biochemical effects might have a similar therapeutic role.

Because AMPA potentiators appear to have antidepressant-like properties (as discussed previously), it is possible that AMPA receptor antagonists could have antimanic effects. The first selective AMPA receptor antagonists to be identified were quinoxalinedone derivatives such as 2,3-dioxo-6-nitro-1,2,3,4-tetrahydrobenzo[*f*]quinoxaline-7-sulfonamide, which is an AMPA receptor antagonist at the glutamate recognition site of the receptor. In contrast to earlier AMPA antagonists that blocked the AMPA recognition site, more recent ones such as 2,3-benzodiazepines (eg, GYKI 52466) block AMPA receptors via an allosteric site on the receptor-channel complex.[134] At present, the GYKI 52466 analog talampanel (GYKI 53773; LY 300164) is undergoing phase III clinical trials. Talampanel, an anticonvulsant, was well tolerated in earlier clinical trials, but sedation may occur, especially with initial dosing (reviewed by Rogawski[135]).

8-Methyl-5-[4-(*N*,*N*-dimethylsulfamoyl)phenyl]-6,7,8,9-tetrahydro-1*H*-pyrrolo[3,2-*h*]-isoquinoline-2,3-dione-3-*O*-(4-hyroxybutyric acid-2yl)oxime (NS 1209; SPD502) is a structurally novel, water-soluble, competitive AMPA receptor antagonist with good central nervous system bioavailability.[136] NS1209 has been well tolerated in phase I/II clinical trials and is being evaluated for the treatment of refractory status epilepticus (reviewed by Rogawski[135]).

Metabotropic Glutamate Receptors (mGluRs)

The mGluRs comprise a family of 8 receptor subtypes (mGluR1 to GluR8) that are classified into 3 groups on the basis of their sequence homology, coupling to second messenger systems, and agonist selectivity. Group I mGluRs (mGluR1 and mGluR5) are coupled to the phospholipase C signal transduction pathway. Group II (mGluR2 and mGluR3) and III (mGluR4 and mGluR6 to mGluR8) receptors are both coupled in an inhibitory manner to the adenylyl cyclase signal transduction pathway, which is generally involved in the regulation of the release of glutamate or other neurotransmitters [eg, γ-aminobutyric acid (GABA)], depending on synaptic localization.[137] The mGluRs are involved in the early phase of memory formation and the mechanism of long-term depression.[138–140] The Group I mGluR5 antagonists 2-methyl-6-[phenylethynyl]-pyridine and [(2-methyl-1,3-thiazol-4-yl)ethynyl]pyridine have shown antidepressant-like activity in the modified FST in rats,[141] in the tail suspension test in mice,[141] and in olfactory bulbectomized rats.[142] Furthermore, 2-methyl-6-[phenylethynyl]-pyridine treatment was recently found to increase the hippocampal mRNA level.[143] The mGluR1 antagonist [3-ethyl-2-methyl-quinolin-6-yl]-(4-methyoxy-cyclohexyl)-methanone methanesulfonate was active in the modified FST in rats and in the tail suspension test in mice.[144]

The mode of antidepressant-like activity of mGluR1 or mGluR5 antagonists is uncertain; some have suggested that inhibitors of mGluR5 might produce a final effect that is similar to that evoked by NMDA antagonists, which are known to display antidepressant-like activity (as discussed previously). It remains unclear whether mGluR5 antagonists will be safe enough to use clinically. Acamprosate is a weak mGluR5 antagonist; studies conducted with it in BPD have been presented at scientific meetings and reportedly did not show efficacy. Several possibilities exist for the lack of improvement with this agent, including its weak affinity for different glutamate receptors and its poor oral absorption. Studies conducted with the non-benzodiazepine anxiolytic fenobam, a potent and selective mGluR5 antagonist, were discontinued because of psychostimulant effects that occurred with its use.[145] More recently, the mGluR5-positive allosteric modulator 3-cyano-*N*-(1,3-diphenyl-1*H*-pyrazol-5-yl)benzamide was found to be a brain penetrant and to reverse amphetamine-induced locomotor activity and amphetamine-induced deficits in prepulse inhibition in rats; these 2 models were thought to be sensitive to antipsychotic treatment.[146] Should this compound result in an antipsychotic drug for clinical use, it is likely that it will be used clinically in patients experiencing a manic episode. Unfortunately, no preclinical data have been published on whether it has an antidepressant profile.

Group II mGluR2 and mGluR2/3 are negatively linked to the adenylyl cyclase signal transduction pathway and decrease glutamate release, especially under conditions of glutamate excess in the synapse; moreover, they regulate glutamate transmission by postsynaptic mechanisms. Group II mGluRs agonists (eg, LY341495) dose-dependently decreased the immobility time of mice in the tail suspension test and reduced immobility time and increased swimming behavior without affecting climbing behavior in rats.[147] Moreover, MGS-0039 has been reported to be effective in the learned helplessness model of depression[148] and to increase cell proliferation in the adult mouse hippocampus.[149] Activation of AMPA receptors has been reported to be responsible at least in part for the antidepressant-like activity of group II mGluR antagonists; the AMPA antagonist 2,3-dioxo-6-nitro-1,2,3,4-tetrahydrobenzo[*f*]quinoxaline-7-sulfonamide blocked the antidepressant-like activity of MGS-0039 in the tail suspension test in mice.[150]

A type 2 mGluR receptor agonist, LY 354740, was recently tested in patients with panic disorder. LY354740 failed to separate from placebo in the primary panic measures, whereas paroxetine did. Although the MADRS scale was collected at baseline and at week 9, no MADRS score was reported for these time points. The compound was generally well tolerated; the most commonly reported adverse events were gastrointestinal complaints (nausea, diarrhea, and stomach pain) and headache. Two of 18 subjects discontinued the compound because of an allergic reaction.[151] A multicenter trial of LY35470 for the treatment of panic disorder was suspended because preclinical studies showed convulsions in mice.[152] No noticeable antipsychotic effects were noted with LY35470, as indicated by its inability to block the increase in Positive and Negative Syndrome Scale scores due to ketamine infusion.[153] Unfortunately, no information is available in depression paradigms.

Pilc and colleagues demonstrated that a selective group III mGluR agonist ([1*S*,3*R*,4S]-1-aminocyclopentane-1,3,4-tricarboxylic acid) and an mGluR8 agonist ([*RS*]-4-phosphonophenylglycine) for this receptor resulted in antidepressant-like effects in the FST in rats.[154] Additional proof for the involvement of group II mGluRs comes from the work of Cryan and colleagues,[155] who showed that mGluR7 knockout in mice produced antidepressant-like effects in the FST and in the tail suspension test in mice. Palucha and colleagues[156] similarly reported the antidepressant-like effects of group III mGluR agonists in the behavioral despair test.

No group III mGluR agonists are yet at the clinic stage for testing. Although there have been no studies with these compounds in animal models of mania, the mGluR5 antagonists and group II mGluR antagonists seem to be very promising compounds, with considerable potential antidepressant-like activity.

HPA Axis

Dysfunction of the HPA axis has been well described in bipolar depression. Hypercortisolemia may be central to the etiopathogenesis of both the depressive symptoms and the neurocognitive deficits observed in BPD. Strategies to better regulate the effects of cortisol, which may potentially restore HPA axis integrity, have been the focus of recent research. The antiglucocorticoid agents studied in the treatment of depression include both cortisol synthesis inhibitors (aminoglutethimide, ketoconazole, and metyrapone) and corticosteroid receptor antagonists (mifepristone),[157] hydrocortisone and dexamethasone,[158,159] and dehydroepiandrosterone (DHEA; reviewed by Quiroz *et al.*[160]). Treatment with glucocorticoid synthesis inhibitors (eg, ketoconazole and metyrapone), the corticotropin releasing factor type 1 (CRF 1) receptor antagonist R-121919, and DHEA has been observed to ameliorate depressive symptoms in patients with unipolar depression (reviewed by Quiroz *et al.*,[160] Jahn *et al.*,[161] Zobel *et al.*,[162] and Schmidt *et al.*[163]). Clinical development of R-121919 was terminated because of its association with abnormal liver function tests.[164] DHEA has also been reported to be associated with a high propensity to induce mania.[165]

The only antiglucocorticoids that have been tested in BPD are ketoconazole and mifepristone. Ketoconazole (up to 800 mg/day) was given as an add-on therapy in 6 depressed patients who had a diagnosis of treatment-resistant BPD.[166] Three patients who received a dose of at least 400 mg/day had substantial reductions in depressive symptoms and no development of manic symptoms; cortisol levels were not lowered in any of the subjects. The significant toxicity risk and drug interactions with ketoconazole preclude its use on a chronic basis for mood disorders.

Only 1 placebo-controlled study of an antiglucocorticoid for bipolar depression has been performed. Mifepristone (RU-486) is a nonselective antagonist of the glucocorticoid receptor that has been reported to have antidepressant and antipsychotic properties in patients with psychotic depression[157] (reviewed by Quiroz *et al.*[160]), although a recent letter to the editor indicates that 2 large phase III studies failed to find significant antipsychotic or antidepressant effects.[167] Animal studies suggest that glucocorticoid receptor numbers are increased rapidly (within hours) after the administration of RU-486, which may restore normal feedback, thus "resetting" the HPA axis. In a recent double-blind, placebo-controlled crossover study, Young and colleagues[168] compared mifepristone (600 mg) to placebo in 20 subjects with bipolar depression. Over the course of the 6-week study, neurocognitive and neuroendocrine function and mood symptoms were measured. The study found not only benefits in depressive symptoms with mifepristone but also benefits in cognitive functioning, specifically in spatial memory. If mifepristone is found to have a beneficial effect in BPD, its use will most likely be limited to acute depressive episodes; long-term treatment would have significant side effects, including the potential for adrenal insufficiency and hepatic injury.[169] In addition, mifepristone has significant antiprogesterone effects, and its chronic use could lead to fatigue, hot flashes, gynecomastia/breast tenderness, and endometrial hyperplasia.[170] A large controlled study with mifepristone in bipolar depression is currently underway (NCT0035912 by Alan Young at the University of British Columbia). Examples of other GR antagonists under development are provided in Table 1.

CRF 1 Receptor Antagonists A number of small-molecule CRF 1R antagonists have been evaluated with in vivo paradigms in animal models to attenuate CRF-induced adrenocorticotropic hormone (ACTH) release.[171] Several classes of CRF 1R inhibitors have been identified (Table 1; see Saunders and Williams[171] and Holmes *et al.*[172] for reviews). In preclinical studies, CRF 1 antagonists diminished CRF-induced ACTH release as well as CRF-induced cyclic adenosine monophosphate production (see Saunders and Williams[171] for a review). In an open-label study, R-121919 reduced anxiety and depressive symptoms in patients with major depression.[162] Some companies have discontinued clinical development of these compounds because of laboratory abnormalities; others are still being developed.

Antalarmin, a novel pyrrolopyrimidine compound, in oral doses of 20 mg/kg in primates significantly reduced CRF-stimulated ACTH release as well as the pituitary-adrenal, sympathetic, and adrenal medullary responses to stress. It also reversed

Table 1. *Examples of Candidate Drugs for the Treatment of Bipolar Disorder.*

System	Drug/Compound*
Dynorphin opioid neuropeptide system	Kappa opioid receptor antagonists: questionable antidepressant effects Kappa opioid receptor agonists: questionable antimanic effects, pentazocine (Talwin; partial agonist for kappa opiate receptor)
Cholinergic system	Anticholinergic: scopolamine (antimuscarinic) and donepezil (cholinesterase inhibitor)
Nicotinic acetylcholine receptor	Cytisine (partial agonist of $\alpha4/\beta2$ and a full agonist at $\alpha3/\beta4$); SIB-1508Y ($\alpha4/\beta2$ agonist); mecamylamine (nicotinic antagonist)
Melatonin and serotonin system	Agomelatine (melatonin MT1 and MT2 agonist and 5-HT2C antagonist)
GSK	GSK inhibitors: zinc, indirubines, maleimides, hymenialdesine, paullones, thiaziazolidones, synthetic phosphorylated peptide, and azole derivatives
PKC	PKC inhibitors: tamoxifen, LY33531, ruboxistaurin, rottlerin, indolocarbazoles, UCN-01, CGP41251, PKC412, bisindolylmaleimides, balanol, indolylindazolylmaleimides, and aprinocarsen
Arachidonic acid metabolism	Celecoxib (COX-2 inhibitors)
Glutamate release and AMPA receptor	Riluzole (inhibitor of glutamate release, enhancer of AMPA trafficking and glutamate synaptic clearance)
NMDA receptor complex	D-Cycloserine (partial agonist glycine site); NMDA antagonist: memantine and ketamine
AMPA receptor potentiators	Benzoylpiperidone (aniracetam), benzoylpyrrolidines (ampakines), arylpropylsulfonames (LY392098, LY451616), and S18986
mGluR	MPEP, MTEP, EMQMCM, and CDPPB (group I mGluR5 antagonists); LY341495 and MGS-0039 (group II mGluR2s); GluR2/3 agonists: ACPT-I and RS-PPG (III mGluR agonists)
Glucocorticoid synthesis	Glucocorticoid synthesis inhibitors: ketoconazole, aminogluthethimide, and metyrapone
GR II receptor	GR II receptor antagonists: mifepristone (RU-486), ORG 34517, ORG 34850, ORG 34116, AL082D06, and cyproterone acetate
CRF 1 receptors	CRF 1R antagonists: peptides (astressin and α-helCRF) and small-molecule nonpeptides (CP-154526, antalarmin, DMP695, DMP696, CRA1000, R-121919, SSR125543, NBI 35965, and NBI 27914)
Bcl-2 enhancer	Pramipexole

Abbreviations: 5-HT2C, 5-hydroxytryptamine (serotonin) receptor 2C; ACPT-I, [1*S*,3*R*,4*S*]-1-aminocyclo-pentane-1,3,4-tricarboxylic acid; AMPA, α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid Bcl-2, B cell lymphoma-2; CDPPN, 3-cyano-*N*-(1,3-diphenyl-1H-pyrazol-5-yl)benzamide; COX, cyclooxygenase; CRF, corticotropin releasing factor; EMQMCM, [3-ethyl-2-methyl-quinolin-6-yl]-(4-methyoxy-cyclohexyl)-methanone methanesulfonate; GC, glucocorticoid; GluR, glutamate receptor; GSK, glycogen synthase kinase; mGluR, metabotropic glutamate receptor; MPEP, 2-methyl-6-[phenylethynyl]-pyridine; MTEP, [(2-methyl-1,3-thiazol-4-yl)ethynyl]pyridine; NMDA, *N*-methyl-D-aspartate receptor; PKC, protein kinase C; RS-PPG, [*RS*]-4-phosphonophenylglycine.

* Drugs/compounds are at different stages of development; some may have been tested for proof of concept rather than for clinical use, and others at this time may not be usable on a long-term basis because of treatment-limiting side effects. Although valproate has plasticity-enhancing characteristics, it remains unclear whether they are due to valproate's ability to inhibit GSK-3 and histone deacetylase.

stress-induced inhibition of exploratory and sexual behaviors.[173] With the chronic stress model in mice, both antalarmin (10 mg/kg) and fluoxetine (10 mg/kg) significantly improved measures of physical state, weight gain, and emotional response in the light-dark test in comparison with stressed, untreated animals.[174] Antalarmin reduced swim-stress–induced ACTH response but did not show antidepressant-like effects in the FST.[175]

CP-154,526, developed by Pfizer, has been evaluated in animal paradigms for anxiety. Like antalarmin, it has high brain-barrier penetrability, decreases synthesis of CRF in the paraventricular nucleus,[176] and shows antidepressant-like properties in the learned helplessness model of depression in rats.[177] SSR125543A, a 2-aminothiazole derivative that displays high affinity for human CRF R1 receptors, has shown efficacy in the FST model and in the chronic mild stress model in rats.[178] In addition, SSR125543A was able to reverse stress-induced suppression of neurogenesis in mice subjected to chronic mild stress,[179] and it has also been shown to reduce aggressive behaviors in male Syrian hamsters in the resident-intruder aggression test.[180] In other studies, CRA 1000, a nonpeptide pyrimidine CRF 1 antagonist, was found to reduce immobility in the learned helplessness paradigm in male Wistar rats.[181] DMP696, developed by Dupont, is a selective, potent, and highly bioavailable nonpeptide CRF 1R antagonist. It has been tested in behavioral models of anxiety and is being tested in behavioral paradigms for depression.[182]

B Cell Lymphoma-2 (Bcl-2) Enhancers As discussed previously, mood disorders are increasingly being found to be associated with problems in neuroplasticity and cellular resilience. Neurotrophic factors are known to promote cell survival largely by suppressing intrinsic, cellular apoptotic machinery. Cell survival appears to occur through the binding of these factors to membrane receptors and regulation of intracellular signal transduction pathways that can control apoptosis, including regulation of the Bcl-2 family of proteins. Bcl-2 is not only neuroprotective; it also exerts neurotrophic effects, promotes neurite sprouting and outgrowth, and promotes axonal regeneration. Recent studies have demonstrated that severe stress exacerbates stroke outcome by suppressing Bcl-2 expression[183]; stressed mice expressed approximately 70% less Bcl-2 mRNA than unstressed mice after ischemia in a stroke model. Enhanced Bcl-2 expression appears to offset the potentially harmful consequences of stress-induced neuronal endangerment, and this suggests that up-regulation of Bcl-2 by pharmacological means may have substantial value in the treatment of mood disorders.

Bcl-2 has traditionally been viewed as a "long-term neuroprotective protein"; however, Bcl-2 is also a major regulator of mitochondrial function, and there is increasing proof of the various roles that mitochondria play in modulating integrated central nervous system function. Thus, mounting evidence suggests that mitochondrial Ca^{2+} sequestration plays a key role in modulating the tone of synaptic plasticity in a variety of neuronal circuits and that regulation of mitochondrial function is likely to play an important role in regulating the synaptic strength of neuronal circuitry mediating complex behaviors. Indeed, it is quite possible that lithium's antidepressant potentiating effects are due to its ability to robustly up-regulate Bcl-2. It is also noteworthy that pramipexole similarly up-regulates Bcl-2 in several brain areas[184] and has been shown to exert antidepressant effects in 2 small, double-blind, placebo-controlled trials in patients with bipolar depression.[185,186] Future studies would have to test whether selective Bcl-2 enhancers without dopamine effects also have antidepressant effects in bipolar depression.

INTRACELLULAR PATHWAYS AND TARGETS WORTHY OF FURTHER STUDY IN BPD

GSK-3

GSK-3 is a serine/threonine kinase that is normally highly active in cells and is deactivated by signals originating from numerous signaling pathways (eg, the Wnt pathway, the phosphoinositide 3-kinase pathway, protein kinase A, and PKC).

Interest in GSK-3 is in line with contemporary theories of mood disorders that point to their association with impairments of neuroplasticity and cellular resilience; neuroimaging findings show regional reductions in brain volume corresponding at the tissue level to decreases in the number, size, and density of neurons and glia precisely in critical circuits purportedly involved in mood disorders. In general, increased activity of GSK-3 is proapoptotic, whereas inhibiting GSK-3 attenuates or prevents apoptosis. Lithium has neurotrophic and neuroprotective properties in rodent and cell-based models, and these effects are clinically suggestive of neuroprotection. Lithium may exert these neuroprotective effects in part by inhibiting GSK-3. More recent preclinical evidence implicates the modulation of GSK-3 in either the direct or downstream mechanism of action of many other mood stabilizer and antidepressant medications currently being prescribed (see Gould and Manji[187]).

Some of the behavioral effects of lithium may be due to inhibition of GSK-3. Pharmacological inhibition of GSK-3 attenuates D-amphetamine hyperlocomotion in rats, which is believed to represent, albeit imperfectly, an animal model of mania.[188] Furthermore, mice overexpressing a constitutively active form of GSK-3β in the brain showed increased locomotor activity as well as decreased habituation in an open field. In support of this notion, we found antidepressant-like behavior in the FST and antimanic-like response to amphetamine following administration of the GSK-3 inhibitor AR-A014418.[188,189] Thus, decreased expression of GSK-3β resulted in attenuation of stimulant-induced locomotion, whereas its increased expression resulted in an endogenously high level of activity.

Proof-of-principle studies with selective GSK inhibitors are urgently needed in BPD to determine the relevance of this target. At present, no blood-brain-penetrant GSK-selective inhibitors are available for human use (Table 1).

PKC Signaling Cascade

PKC is highly enriched, has a heterogeneous distribution in the brain, and plays an important role in regulating neuronal excitability, neurotransmitter release, and long-term alterations in gene expression and plasticity. A considerable amount of biochemical data support the potential involvement of PKC and its substrates in bipolar patients and changes in PKC signaling pathways after treatment with lithium or valproate.[190–195] These findings provide ample evidence that the PKC signaling pathway is clearly a target for the actions of 2 structurally highly dissimilar antimanic agents–lithium and valproate–and provide the impetus to test a PKC inhibitor in mania. Although best known for its anti-estrogenic properties, tamoxifen is also a potent PKC inhibitor, especially at high concentrations. In animals, tamoxifen has no effect on the resident-intruder test in males (an "animal model of mania") or in the FST in mice[196] but does reduce amphetamine-induced hyperactivity in a large open field and amphetamine-induced phosphorylation of growth-associated protein 43.[197]

In humans, tamoxifen was found to have significant antimanic effects. In a single-blind study, tamoxifen treatment significantly decreased manic symptoms in 5 of 7 patients enrolled in the initial trial.[198] In another recently completed double-blind, placebo-controlled trial with tamoxifen in patients with bipolar mania, tamoxifen exhibited significant antimanic effects in doses as high as 140 mg/day as early as day 5 and throughout the 3 weeks of the trial.[199] The antimanic effect of tamoxifen was not the result of sedation. There was no increased risk of depression with tamoxifen compared to placebo in this short-term study. Whether tamoxifen is associated with an increased risk of depression if used on a long-term basis in patients with mood disorders is unknown.

Other studies conducted with tamoxifen also confirm the relevance of PKC inhibition in antimanic agents.[200,201] Other drugs with PKC inhibitory effects include omega-3 fatty acids and verapamil, but these effects are very weak and perhaps in part explain why they have not been found to be consistently effective in BPD. With respect to the selectivity of tamoxifen's effects on PKC, it is important to re-emphasize that tamoxifen is also an anti-estrogen. It is possible that some of the antimanic effects seen with tamoxifen are attributable to estrogen receptor antagonism (see Goldstein[201]). To our knowledge, the PKC line of research provides evidence for the first time in BPD drug development of a direct molecular target whose inhibition results in antimanic effects in humans. The role of PKC inhibition in bipolar depression or in long-term maintenance treatment is unknown at this time. Large controlled studies with selective PKC inhibitors in acute bipolar mania are warranted.

AA Cascade

Accumulating data suggest that an inflammatory process is involved in the pathophysiology of mood disorders. The enzymes that regulate the brain AA cascade have been implicated in BPD. AA functions as a key intermediary of second messenger pathways within the brain. It is released from membrane phospholipids via receptor/G protein–initiated activation of phospholipase A2. This action results in the release of AA from the cellular membrane and cyclooxygenase (COX)-mediated production of eicosanoid metabolites such as prostaglandins and thromboxanes. These metabolites mediate many subsequent intracellular and transynaptic responses.

Chronic treatment with lithium and valproate in rats selectively reduces the turnover rate in brain phospholipids of AA, which are believed to be hyperactive in mania.[202] In addition, lithium downregulates the gene expression and protein levels of an AA-specific phospholipase as well as the protein levels of COX-2. Valproate decreases turnover of AA, protein levels of COX-1 and COX-2,[203] and frontal cortex COX-2 mRNA.[204] COX-2 also protects against neurotoxicity promoted by excessive concentrations of glutamate. These findings suggest that the effects of mood-stabilizer therapies on cell membranes and specifically on AA turnover might be relevant to the mechanism of action of lithium and valproate.

Additional evidence for the involvement of the AA signaling pathway in BPD comes from other preclinical and clinical studies. Administration of the nonselective COX inhibitors indomethacin and piroxicam in rats prevented amphetamine-stimulated locomotor activity and blocked cocaine sensitization (both are rodent models of mania). Moreover, inhibition of COX-2 with NS-398 attenuated restraint stress–induced oxidative changes (a model of depression). In olfactory bulbectomized rats, the COX-2 inhibitor celecoxib showed antidepressant-like properties.[205] In humans, celecoxib was also found to have antidepressant properties. In a 6-week, double-blind, placebo-controlled trial, Muller and colleagues[206] found that celecoxib (400 mg/day), when added to reboxetine in patients with major (unipolar) depression, produced significant antidepressant effects in comparison with placebo. Recently, a 6-week, double-blind, placebo-controlled trial found that celecoxib (400 mg/day) was effective in patients with bipolar I or II depression when added to ongoing mood stabilizer treatment; however, add-on celecoxib was more effective than placebo only at week 1, not at the study endpoint.[207] Notably, it remains unclear how brain-penetrant celecoxib actually is. In addition, it is also presently unclear whether directly targeting COX-2 is worthwhile because selective COX-2 inhibitors may be associated with an increased risk of adverse cardiovascular outcomes.[208]

OTHER DRUGS BEING TESTED IN BPD

Modafinil is approved as a wake-promoting agent for the treatment of excessive daytime sleepiness in narcoleptic patients.[209] The precise mechanism of action of modafinil is unknown, but it is believed to operate on multiple systems, including glutamate, GABA, hypocretin, and, to a lesser degree, dopaminergic and noradrenergic systems. More specifically, it is known to activate noradrenergic $\alpha 1$ receptors, to increase phosphorylation of mitogen-activated protein kinase in cultured mouse cells, to decrease GABA release in the nucleus accumbens of rats, to weakly increase dopamine in the nucleus accumbens secondary to decreased GABA,[210] to increase the release of glutamate in the hippocampal formation and ventromedial and ventrolateral areas of the thalamus, and to activate hypocretin-secreting neurons in the perifornical area. Modafinil is being increasingly used in mood disorders with apparent beneficial effects. In a 6-week, randomized, double-blind, placebo-controlled evaluation of modafinil (mean dose: 177 mg) in subjects with bipolar I or II depression who were inadequately responsive to mood stabilization with or without adjunctive antidepressant therapy ($n = 87$), there was greater baseline-to-endpoint change in modafinil versus placebo and from week 2 on between the groups.[211] No manic switches were reported. In another study, Frye and colleagues (unpublished data reported by Post *et al.*[212]) compared the adjunctive use of modafinil (100 or 200 mg in the morning for 3 weeks) with placebo in the treatment of patients with BPD and residual depressive symptoms, fatigue, or both. Modafinil was significantly more effective than placebo on a variety of measures, including the baseline-to-endpoint change on the Inventory for Depressive Symptoms, percentage response rate, remission rate, and clinical global impression improvement. Modafinil was not significantly associated with a treatment-emergent manic episode. One case report described a manic switch in a patient with treatment-resistant bipolar depression treated with modafinil,[213] and 2 cases of modafinil-induced irritability and aggression in 2 bipolar patients have been described.[214]

A proprietary formulation of the nucleoside uridine, Uridine RG2417 (Repligen Corp.), a biological compound essential for the synthesis of DNA and RNA, is currently in development for the treatment of neuropsychiatric disorders and neurodegenerative diseases. A previous study suggested that RG2133, the prodrug of RG2419, had antidepressant-like effects in an animal model of depression. RG2417 was recently found to be efficacious in a phase 2a multisite study in bipolar depression. This was a multicenter study in which 84 patients received either RG2417 or placebo twice a day. Over the 6-week treatment period, patients receiving RG2417 demonstrated a statistically significant improvement in the symptoms of depression in comparison with those patients receiving placebo on the MADRS ($P = 0.03$) and a strong trend toward improvement on the Clinical Global Impression of Change in Bipolar Disorder Scale ($P = 0.06$). This study was conducted under a development agreement with the Stanley Medical Research Institute (study ID number NCT00322764; http://www.medicalnewstoday.com/articles/88213.php).

A neuronal L-type calcium channel modulator (MEM 1003, Memory Pharmaceuticals) was found to lack efficacy in acute mania in a phase 2a multisite study (study ID number NCT00374920;

http://www.genengnews.com/news/bnitem.aspx?name=13780779). In that study, approximately 80 subjects were randomized to receive MEM 1003 (120–360 mg/day) or placebo for 21 days.

CONCLUSIONS

By the criteria specified for this review article, a number of candidate targets were found that could result in putative treatments for BPD. These include (1) the purinergic system, (2) the dynorphin opioid neuropeptide system, (3) the cholinergic system (muscarinic and nicotinic systems), (4) the melatonin and serotonin (5-HT2C receptor) system, (5) the glutamatergic system, and (6) the HPA axis as well as (7) GSK-3 protein, (8) PKC, (9) the AA cascade, and (10) other systems.

Drug development for BPD may occur through 1 of 2 approaches.[93] The first is through an understanding of the therapeutically relevant biochemical targets of currently effective medications; examples of this approach have been reviewed here and include the common targets of lithium and valproate, which are PKC and GSK-3β. The investigation of PKC illustrates how this approach could lead to the development of therapeutics for BPD. In the first phase, the shared molecular targets of lithium and valproate were identified: in this case, PKC. Next, the therapeutic relevance of this finding was established in preclinical clinical studies and in biochemical studies in patients suffering from this disorder. Finally, and most notably, the relevance of this target was tested in humans; the end result was that the PKC inhibitor tamoxifen was found to result in antimanic effects in patients with bipolar mania.

The second path for drug development results from our understanding of the cellular and molecular underpinnings of severe mood disorders and the manner in which they are associated with regional impairments of structural plasticity and cellular resiliency. Newer "plasticity-enhancing" strategies that may be useful in the treatment of mood disorders include inhibitors of glutamate release, NMDA antagonists, and glucocorticoid receptor antagonists.

Finally, several points merit further discussion. First, the drugs/drug mechanisms reviewed here are potentially worthy of study in 1 pole of the illness, but it remains unclear whether they would be equally effective for the other pole of the illness or in maintenance treatment. Second, the proof-of-concept studies reviewed here are based on very small sample sizes, and additional study of these drugs is necessary in larger controlled trials before they can be generalized to current clinical practice. Third, when one is deciding on a drug mechanism or target to pursue in drug development for BPD, consideration should be giving to the following set criteria. Not doing so increases the risk of developing drugs that are not likely to succeed. For example, many lessons were learned when the GABAergic agents failed to demonstrate efficacy in mania; they were chosen and developed primarily on the presumptive drug mechanism alone (reviewed by Goodnick[215]).

Enriched criteria for establishing target validation for further development in BPD have been proposed.[130,191] These include (1) corroboration of a target at the protein and functional level; (2) observation with chemically dissimilar but clinically effective agents; (3) occurrence at a dose/plasma level and time frame consistent with clinical therapeutic effect; (4) localization to brain regions implicated in the neurobiology of the disorder under consideration; (5) when known, relevance to known pathophysiology; and (6) when possible, tethering to human genetic findings. Indeed, such a strategy was recently used in the development of the PKC inhibitors.[216]

ACKNOWLEDGMENT

The authors acknowledge the support of the Intramural Research Program of the National Institute of Mental Heath, the Stanley Medical Research Institute, and the National Alliance for Research on Schizophrenia And Depression. Carlos A. Zarate Jr and Husseini K. Manji are listed among the inventors on a patent application submitted for the use of ketamine in depression. They have assigned their rights on the patent to the US government.

DISCLOSURES

There is nothing to report.

REFERENCES

1. Merikangas KR, Akiskal HS, Angst J, et al. Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. *Arch Gen Psychiatry* 2007; 64: 543–552.
2. Gitlin M. Treatment-resistant bipolar disorder. *Mol Psychiatry* 2006; 11: 227–240.
3. Judd LL, Akiskal HS, Schettler PJ, et al. The long-term natural history of the weekly symptomatic status of bipolar I disorder. *Arch Gen Psychiatry* 2002; 59: 530–537.

4. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry* 2006; 163: 1905–1917.
5. Trivedi MH, Rush AJ, Wisniewski SR, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry* 2006; 163: 28–40.
6. Sachs GS, Nierenberg AA, Calabrese JR, et al. Effectiveness of adjunctive antidepressant treatment for bipolar depression. *N Engl J Med* 2007; 356: 1711–1722.
7. Manji HK, Zarate CA. Molecular and cellular mechanisms underlying mood stabilization in bipolar disorder: implications for the development of improved therapeutics. *Mol Psychiatry* 2002; 7(suppl 1): S1–S7.
8. Einat H, Manji HK. Cellular plasticity cascades: genes-to-behavior pathways in animal models of bipolar disorder. *Biol Psychiatry* 2006; 59: 1160–1171.
9. Ban TA. The role of serendipity in drug discovery. *Dialogues Clin Neurosci* 2006; 8: 335–344.
10. Cade JF. Lithium salts in the treatment of psychotic excitement. *Med J Aust* 1949; 2: 349–352.
11. Anumonye A, Reading HW, Knight F, Ashcroft GW. Uric-acid metabolism in manic-depressive illness and during lithium therapy. *Lancet* 1968; 1: 1290–1293.
12. Machado-Vieira R, Lara DR, Souza DO, Kapczinski F. Purinergic dysfunction in mania: an integrative model. *Med Hypotheses* 2002; 58: 297–304.
13. Barden N, Harvey M, Gagne B, et al. Analysis of single nucleotide polymorphisms in genes in the chromosome 12Q24.31 region points to P2RX7 as a susceptibility gene to bipolar affective disorder. *Am J Med Genet B Neuropsychiatr Genet* 2006; 141: 374–382.
14. Lucae S, Salyakina D, Barden N, et al. P2RX7, a gene coding for a purinergic ligand-gated ion channel, is associated with major depressive disorder. *Hum Mol Genet* 2006; 15: 2438–2445.
15. Ogawa N, Ueki H. Secondary mania caused by caffeine. *Gen Hosp Psychiatry* 2003; 25: 138–139.
16. Lara DR, Dall'Igna OP, Ghisolfi ES, Brunstein MG. Involvement of adenosine in the neurobiology of schizophrenia and its therapeutic implications. *Prog Neuropsychopharmacol Biol Psychiatry* 2006; 30: 617–629.
17. Kelley WN. Effects of drugs on uric acid in man. *Annu Rev Pharmacol* 1975; 15: 327–350.
18. Machado-Vieira R, Lara DR, Souza DO, Kapczinski F. Therapeutic efficacy of allopurinol in mania associated with hyperuricemia. *J Clin Psychopharmacol* 2001; 21: 621–622.
19. Akhondzadeh S, Milajerdi MR, Amini H, Tehrani-Doost M. Allopurinol as an adjunct to lithium and haloperidol for treatment of patients with acute mania: a double-blind, randomized, placebo-controlled trial. *Bipolar Disord* 2006; 8: 485–489.
20. Machado-Vieira R, Soares JC, Lara DR, et al. A double-blind, randomized, placebo-controlled 4-week study on the efficacy and safety of the purinergic agents allopurinol and dipyridamole adjunctive to lithium in acute bipolar mania. *J Clin Psychiatry* 2008.; Available at: http://www.psychiatrist.com.
21. Hurd YL. Subjects with major depression or bipolar disorder show reduction of prodynorphin mRNA expression in discrete nuclei of the amygdaloid complex. *Mol Psychiatry* 2002; 7: 75–81.
22. Rees DC. Chemical structures and biological activities of non-peptide selective kappa opioid ligands. *Prog Med Chem* 1992; 29: 109–139.
23. Kuzmin AV, Gerrits MA, Van Ree JM. Kappa-opioid receptor blockade with nor-binaltorphimine modulates cocaine self-administration in drug-naive rats. *Eur J Pharmacol* 1998; 358: 197–202.
24. Jewett DC, Grace MK, Jones RM, et al. The kappa-opioid antagonist GNTI reduces U50,488-, DAMGO-, and deprivation-induced feeding, but not butorphanol- and neuropeptide Y-induced feeding in rats. *Brain Res* 2001; 909: 75–80.
25. Roth BL, Baner K, Westkaemper R, et al. Salvinorin A: a potent naturally occurring nonnitrogenous kappa opioid selective agonist. *Proc Natl Acad Sci U S A* 2002; 99: 11 934–11 939.
26. Carlezon WA Jr, Beguin C, DiNieri JA, et al. Depressive-like effects of the kappa-opioid receptor agonist salvinorin A on behavior and neurochemistry in rats. *J Pharmacol Exp Ther* 2006; 316: 440–447.
27. Barber A, Gottschlich R. Novel developments with selective, non-peptidic kappa-opioid receptor agonists. *Expert Opin Investig Drugs* 1997; 6: 1351–1368.
28. Mague SD, Pliakas AM, Todtenkopf MS, et al. Antidepressant-like effects of kappa-opioid receptor antagonists in the forced swim test in rats. *J Pharmacol Exp Ther* 2003; 305: 323–330.
29. Reindl JD, Rowan K, Carey AN, et al. Antidepressant-like effects of the novel kappa opioid antagonist MCL-144B in the forced-swim test. *Pharmacology* 2008; 81: 229–235.
30. Rimoy GH, Wright DM, Bhaskar NK, Rubin PC. The cardiovascular and central nervous system effects in the human of U-62 066E. A selective opioid receptor agonist. *Eur J Clin Pharmacol* 1994; 46: 203–207.
31. Walsh SL, Strain EC, Abreu ME, Bigelow GE. Enadoline, a selective kappa opioid agonist: comparison with butorphanol and hydromorphone in humans. *Psychopharmacology (Berl)* 2001; 157: 151–162.
32. Cohen BM, Murphy B. The effects of pentazocine, a kappa agonist, in patients with mania. *Int J Neuropsychopharmacol* 2008; 11: 243–247.
33. Challoner KR, McCarron MM, Newton EJ. Pentazocine (Talwin) intoxication: report of 57 cases. *J Emerg Med* 1990; 8: 67–74.
34. Belleville JP, Dorey F, Bellville JW. Effects of nefopam on visual tracking. *Clin Pharmacol Ther* 1979; 26: 457–463.
35. Valdes LJ III. Salvia divinorum and the unique diterpene hallucinogen, Salvinorin (divinorin) A. *J Psychoactive Drugs* 1994; 26: 277–283.
36. Sheffler DJ, Roth BL Salvinorin A: the "magic mint" hallucinogen finds a molecular target in the kappa opioid receptor. *Trends Pharmacol Sci* 2003; 24: 107–109.
37. Janowsky DS, El-Yousef MK, Davis JM, Sekerke HJ. A cholinergic-adrenergic hypothesis of mania and depression. *Lancet* 1972; 2: 632–635.
38. Shi J, Hattori E, Zou H, et al. No evidence for association between 19 cholinergic genes and bipolar

disorder. *Am J Med Genet B Neuropsychiatr Genet* 2007; 144: 715–723.

39. Overstreet DH. The Flinders sensitive line rats: a genetic animal model of depression. *Neurosci Biobehav Rev* 1993; 17: 51–68.
40. Dilsaver SC. Pathophysiology of "cholinoceptor supersensitivity" in affective disorders. *Biol Psychiatry* 1986; 21: 813–829.
41. Sokolski KN, DeMet EM. Cholinergic sensitivity predicts severity of mania. *Psychiatry Res* 2000; 95: 195–200.
42. Cannon DM, Carson RE, Nugent AC, et al. Reduced muscarinic type 2 receptor binding in subjects with bipolar disorder. *Arch Gen Psychiatry* 2006; 63: 741–747.
43. Davis KL, Berger PA, Hollister LE, Defraites E. Physostigmine in mania. *Arch Gen Psychiatry* 1978; 35: 119–122.
44. Khouzam HR, Kissmeyer PM. Physostigmine temporarily and dramatically reversing acute mania. *Gen Hosp Psychiatry* 1996; 18: 203–204.
45. Burt T, Sachs GS, Demopulos C. Donepezil in treatment-resistant bipolar disorder. *Biol Psychiatry* 1999; 45: 959–964.
46. Eden Evins A, Demopulos C, Nierenberg A, et al. A double-blind, placebo-controlled trial of adjunctive donepezil in treatment-resistant mania. *Bipolar Disord* 2006; 8: 75–80.
47. Kasper S, Moises HW, Beckmann H. The anticholinergic biperiden in depressive disorders. *Pharmacopsychiatria* 1981; 14: 195–198.
48. Furey ML, Drevets WC. Antidepressant efficacy of the antimuscarinic drug scopolamine: a randomized, placebo-controlled clinical trial. *Arch Gen Psychiatry* 2006; 63: 1121–1129.
49. Jellinek T. Mood elevating effect of trihexyphenidyl and biperiden in individuals taking antipsychotic medication. *Dis Nerv Syst* 1977; 38: 353–355.
50. Jones CK, Shannon HE. Effects of scopolamine in comparison with apomorphine and phencyclidine on prepulse inhibition in rats. *Eur J Pharmacol* 2000; 391: 105–112.
51. Nashmi R, Lester HA. CNS localization of neuronal nicotinic receptors. *J Mol Neurosci* 2006; 30: 181–184.
52. Lukas RJ, Changeux JP, Le Novere N, et al. International Union of Pharmacology. XX. Current status of the nomenclature for nicotinic acetylcholine receptors and their subunits. *Pharmacol Rev* 1999; 51: 397–401.
53. Glassman AH. Cigarette smoking: implications for psychiatric illness. *Am J Psychiatry* 1993; 150: 546–553.
54. Epping-Jordan MP, Watkins SS, Koob GF, Markou A. Dramatic decreases in brain reward function during nicotine withdrawal. *Nature* 1998; 393: 76–89.
55. Tizabi Y, Overstreet DH, Rezvani AH, et al. Antidepressant effects of nicotine in an animal model of depression. *Psychopharmacology (Berl)* 1999; 142: 193–199.
56. Semba J, Mataki C, Yamada S, et al. Antidepressant-like effects of chronic nicotine on learned helplessness paradigm in rats. *Biol Psychiatry* 1998; 43: 389–391.
57. Mineur YS, Somenzi O, Picciotto MR. Cytisine, a partial agonist of high-affinity nicotinic acetylcholine receptors, has antidepressant-like properties in male C57BL/6J mice. *Neuropharmacology* 2007; 52: 1256–1262.
58. De Luca V, Likhodi O, Van Tol HH, et al. Regulation of alpha7-nicotinic receptor subunit and alpha7-like gene expression in the prefrontal cortex of patients with bipolar disorder and schizophrenia. *Acta Psychiatr Scand* 2006; 114: 211–215.
59. McClernon FJ, Hiott FB, Westman EC, et al. Transdermal nicotine attenuates depression symptoms in nonsmokers: a double-blind, placebo-controlled trial. *Psychopharmacology (Berl)* 2006; 189: 125–133.
60. Ferguson SM, Brodkin JD, Lloyd GK, Menzaghi F. Antidepressant-like effects of the subtype-selective nicotinic acetylcholine receptor agonist, SIB-1508Y, in the learned helplessness rat model of depression. *Psychopharmacology (Berl)* 2000; 152: 295–303.
61. Rabenstein RL, Caldarone BJ, Picciotto MR. The nicotinic antagonist mecamylamine has antidepressant-like effects in wild-type but not beta2- or alpha7-nicotinic acetylcholine receptor subunit knockout mice. *Psychopharmacology (Berl)* 2006; 189: 395–401.
62. Shytle RD, Silver AA, Sanberg PR. Comorbid bipolar disorder in Tourette's syndrome responds to the nicotinic receptor antagonist mecamylamine (Inversine). *Biol Psychiatry* 2000; 48: 1028–1031.
63. Shytle RD, Silver AA, Sheehan KH, et al. Neuronal nicotinic receptor inhibition for treating mood disorders: preliminary controlled evidence with mecamylamine. *Depress Anxiety* 2002; 16: 89–92.
64. Miller DK, Segert IL. Mecamylamine attenuates ephedrine-induced hyperactivity in rats. *Pharmacol Biochem Behav* 2005; 81: 165–169.
65. Tizabi Y, Copeland RL Jr, Brus R, Kostrzewa RM. Nicotine blocks quinpirole-induced behavior in rats: psychiatric implications. *Psychopharmacology (Berl)* 1999; 145: 433–441.
66. Lewy AJ, Wehr TA, Goodwin FK, et al. Manic-depressive patients may be supersensitive to light. *Lancet* 1981; 1: 383–384.
67. Lewy AJ, Nurnberger JI Jr, Wehr TA, et al. Supersensitivity to light: possible trait marker for manic-depressive illness. *Am J Psychiatry* 1985; 142: 725–727.
68. Hallam KT, Olver JS, Norman TR. Melatonin sensitivity to light in monozygotic twins discordant for bipolar I disorder. *Aust N Z J Psychiatry* 2005; 39: 947.
69. Nurnberger JI Jr, Adkins S, Lahiri DK, et al. Melatonin suppression by light in euthymic bipolar and unipolar patients. *Arch Gen Psychiatry* 2000; 57: 572–579.
70. Thomson PA, Wray NR, Thomson AM, et al. Sex-specific association between bipolar affective disorder in women and GPR50, an X-linked orphan G protein-coupled receptor. *Mol Psychiatry* 2005; 10: 470–478.
71. Alaerts M, Venken T, Lenaerts AS, et al. Lack of association of an insertion/deletion polymorphism in the G protein-coupled receptor 50 with bipolar disorder in a Northern Swedish population. *Psychiatr Genet* 2006; 16: 235–236.
72. Hallam KT, Olver JS, Horgan JE, et al. Low doses of lithium carbonate reduce melatonin light sensitivity in healthy volunteers. *Int J Neuropsychopharmacol* 2005; 8: 255–259.
73. Hallam KT, Olver JS, Norman TR. Effect of sodium valproate on nocturnal melatonin sensitivity to light in

healthy volunteers. *Neuropsychopharmacology* 2005; 30: 1400–1404.

74. Leibenluft E, Feldman-Naim S, Turner EH, et al. Effects of exogenous melatonin administration and withdrawal in five patients with rapid-cycling bipolar disorder. *J Clin Psychiatry* 1997; 58: 383–388.
75. Bersani G, Garavini A. Melatonin add-on in manic patients with treatment resistant insomnia. *Prog Neuropsychopharmacol Biol Psychiatry* 2000; 24: 185–191.
76. Van Oekelen D, Luyten WH, Leysen JE. 5-HT2A and 5-HT2C receptors and their atypical regulation properties. *Life Sci* 2003; 72: 2429–2449.
77. Banasr M, Soumier A, Hery M, et al. Agomelatine, a new antidepressant, induces regional changes in hippocampal neurogenesis. *Biol Psychiatry* 2006; 59: 1087–1096.
78. Nagayama H. Chronic administration of imipramine and lithium changes the phase-angle relationship between the activity and core body temperature circadian rhythms in rats. *Chronobiol Int* 1996; 13: 251–259.
79. Armstrong SM, McNulty OM, Guardiola-Lemaitre B, Redman JR. Successful use of S20098 and melatonin in an animal model of delayed sleep-phase syndrome (DSPS). *Pharmacol Biochem Behav* 1993; 46: 45–49.
80. Redman JR, Francis AJ. Entrainment of rat circadian rhythms by the melatonin agonist S-20 098 requires intact suprachiasmatic nuclei but not the pineal. *J Biol Rhythms* 1998; 13: 39–51.
81. Papp M, Gruca P, Boyer PA, Mocaer E. Effect of agomelatine in the chronic mild stress model of depression in the rat. *Neuropsychopharmacology* 2003; 28: 694–703.
82. Bertaina-Anglade V, La Rochelle CD, Boyer PA, Mocaer E. Antidepressant-like effects of agomelatine (S 20 098) in the learned helplessness model. *Behav Pharmacol* 2006; 17: 703–713.
83. Millan MJ, Brocco M, Gobert A, Dekeyne A. Anxiolytic properties of agomelatine, an antidepressant with melatoninergic and serotonergic properties: role of 5-HT2C receptor blockade. *Psychopharmacology (Berl)* 2005; 177: 448–458.
84. Loo H, Hale A D'Haenen H. Determination of the dose of agomelatine, a melatoninergic agonist and selective 5-HT(2C) antagonist, in the treatment of major depressive disorder: a placebo-controlled dose range study. *Int Clin Psychopharmacol* 2002; 17: 239–247.
85. Kennedy SH, Emsley R. Placebo-controlled trial of agomelatine in the treatment of major depressive disorder. *Eur Neuropsychopharmacol* 2006; 16: 93–100.
86. Montgomery SA, Kasper S. Severe depression and antidepressants: focus on a pooled analysis of placebo-controlled studies on agomelatine. *Int Clin Psychopharmacol* 2007; 22: 283–291.
87. Kupfer DJ. Depression and associated sleep disturbances: patient benefits with agomelatine. *Eur Neuropsychopharmacol* 2006; 16(suppl 5): S639–S643.
88. Montgomery SA. Major depressive disorders: clinical efficacy and tolerability of agomelatine, a new melatonergic agonist. *Eur Neuropsychopharmacol* 2006; 16(suppl 5): S633–S638.
89. Calabrese JR, Guelfi JD, Perdrizet-Chevallier C. Agomelatine adjunctive therapy for acute bipolar depression: preliminary open data. *Bipolar Disord* 2007; 9: 628–635.
90. Montgomery SA, Kennedy SH, Burrows GD, et al. Absence of discontinuation symptoms with agomelatine and occurrence of discontinuation symptoms with paroxetine: a randomized, double-blind, placebo-controlled discontinuation study. *Int Clin Psychopharmacol* 2004; 19: 271–280.
91. Zarate CA, Quiroz J, Payne J, Manji HK. Modulators of the glutamatergic system: implications for the development of improved therapeutics in mood disorders. *Psychopharmacol Bull* 2002; 36: 35–83.
92. Sanacora G, Rothman DL, Mason G, Krystal JH. Clinical studies implementing glutamate neurotransmission in mood disorders. *Ann N Y Acad Sci* 2003; 1003: 292–308.
93. Zarate CA Jr, Singh J, Manji HK. Cellular plasticity cascades: targets for the development of novel therapeutics for bipolar disorder. *Biol Psychiatry* 2006; 59: 1006–1020.
94. Bannerman DM, Good MA, Butcher SP, et al. Distinct components of spatial learning revealed by prior training and NMDA receptor blockade. *Nature* 1995; 378: 182–186.
95. Collingridge GL. Long-term potentiation. A question of reliability. *Nature* 1994; 371: 652–653.
96. Collingridge GL, Bliss TV. Memories of NMDA receptors and LTP. *Trends Neurosci* 1995; 18: 54–56.
97. Watkins J, Collingridge G. Phenylglycine derivatives as antagonists of metabotropic glutamate receptors. *Trends Pharmacol Sci* 1994; 15: 333–342.
98. Zarate CA Jr, Singh JB, Carlson PJ, et al. A randomized trial of an N-methyl-D-aspartate antagonist in treatment-resistant major depression. *Arch Gen Psychiatry* 2006; 63: 856–864.
99. Du J, Suzuki K, Wei Y, et al. The anticonvulsants lamotrigine, riluzole, and valproate differentially regulate AMPA receptor membrane localization: relationship to clinical effects in mood disorders. *Neuropsychopharmacology* 2007; 32: 793–802.
100. Frizzo ME, Dall'Onder LP, Dalcin KB, Souza DO. Riluzole enhances glutamate uptake in rat astrocyte cultures. *Cell Mol Neurobiol* 2004; 24: 123–128.
101. Mizuta I, Ohta M, Ohta K, et al. Riluzole stimulates nerve growth factor, brain-derived neurotrophic factor and glial cell line-derived neurotrophic factor synthesis in cultured mouse astrocytes. *Neurosci Lett* 2001; 310: 117–120.
102. Zarate CA Jr, Payne JL, Quiroz J, et al. An open-label trial of riluzole in patients with treatment-resistant major depression. *Am J Psychiatry* 2004; 161: 171–174.
103. Sanacora G, Kendell SF, Levin Y, et al. Preliminary evidence of riluzole efficacy in antidepressant-treated patients with residual depressive symptoms. *Biol Psychiatry* 2007; 61: 822–825.
104. Zarate CA Jr, Quiroz JA, Singh JB, et al. An open-label trial of the glutamate-modulating agent riluzole in combination with lithium for the treatment of bipolar depression. *Biol Psychiatry* 2005; 57: 430–432.
105. Lourenco Da Silva A, Hoffmann A, Dietrich MO, et al. Effect of riluzole on MK-801 and amphetamine-induced hyperlocomotion. *Neuropsychobiology* 2003; 48: 27–30.

106. Cameron HA, McEwen BS, Gould E. Regulation of adult neurogenesis by excitatory input and NMDA receptor activation in the dentate gyrus. *J Neurosci* 1995; 15: 4687–4692.
107. Dall'Olio R, Rimondini R, Gandolfi O. The NMDA positive modulator D-cycloserine inhibits dopamine-mediated behaviors in the rat. *Neuropharmacology* 1994; 33: 55–59.
108. McAllister KH. D-cycloserine enhances social behaviour in individually-housed mice in the resident-intruder test. *Psychopharmacology (Berl)* 1994; 116: 317–325.
109. Heresco-Levy U, Javitt DC, Gelfin Y, et al. Controlled trial of D-cycloserine adjuvant therapy for treatment-resistant major depressive disorder. *J Affect Disord* 2006; 93: 239–243.
110. Lopes T, Neubauer P, Boje KM. Chronic administration of NMDA glycine partial agonists induces tolerance in the Porsolt swim test. *Pharmacol Biochem Behav* 1997; 58: 1059–1064.
111. Chen HS, Wang YF, Rayudu PV, et al. Neuroprotective concentrations of the N-methyl-D-aspartate open-channel blocker memantine are effective without cytoplasmic vacuolation following post-ischemic administration and do not block maze learning or long-term potentiation. *Neuroscience* 1998; 86: 1121–1132.
112. Chen HS, Lipton SA. Mechanism of memantine block of NMDA-activated channels in rat retinal ganglion cells: uncompetitive antagonism. *J Physiol* 1997; 499(pt 1): 27–46.
113. Moryl E, Danysz W, Quack G. Potential antidepressive properties of amantadine, memantine and bifemelane. *Pharmacol Toxicol* 1993; 72: 394–397.
114. Rogoz Z, Skuza G, Maj J, Danysz W. Synergistic effect of uncompetitive NMDA receptor antagonists and antidepressant drugs in the forced swimming test in rats. *Neuropharmacology* 2002; 42: 1024–1030.
115. Zarate CA Jr, Singh JB, Quiroz JA, et al. A double-blind, placebo-controlled study of memantine in the treatment of major depression. *Am J Psychiatry* 2006; 163: 153–155.
116. Ferguson JM, Shingleton RN. An open-label, flexible-dose study of memantine in major depressive disorder. *Clin Neuropharmacol* 2007; 30: 136–144.
117. Teng CT, Demetrio FN. Memantine may acutely improve cognition and have a mood stabilizing effect in treatment-resistant bipolar disorder. *Rev Bras Psiquiatr* 2006; 28: 252–254.
118. Owley T, Salt J, Guter S, et al. A prospective, open-label trial of memantine in the treatment of cognitive, behavioral, and memory dysfunction in pervasive developmental disorders. *J Child Adolesc Psychopharmacol* 2006; 16: 517–524.
119. Wiley JL, Harvey SA, Balster RL, Nicholson KL. Affinity and specificity of N-methyl-D-aspartate channel blockers affect their ability to disrupt prepulse inhibition of acoustic startle in rats. *Psychopharmacology (Berl)* 2003; 165: 378–385.
120. Loscher W, Honack D. High doses of memantine (1-amino-3,5-dimethyladamantane) induce seizures in kindled but not in non-kindled rats. *Naunyn Schmiedebergs Arch Pharmacol* 1990; 341: 476–481.
121. Berman RM, Cappiello A, Anand A, et al. Antidepressant effects of ketamine in depressed patients. *Biol Psychiatry* 2000; 47: 351–354.
122. Borges K, Dingledine R. AMPA receptors: molecular and functional diversity. *Prog Brain Res* 1998; 116: 153–170.
123. Bleakman D, Lodge D. Neuropharmacology of AMPA and kainate receptors. *Neuropharmacology* 1998; 37: 1187–1204.
124. Black MD. Therapeutic potential of positive AMPA modulators and their relationship to AMPA receptor subunits. A review of preclinical data. *Psychopharmacology (Berl)* 2005; 179: 154–163.
125. Martinez-Turrillas R, Frechilla D, Del Rio J. Chronic antidepressant treatment increases the membrane expression of AMPA receptors in rat hippocampus. *Neuropharmacology* 2002; 43: 1230–1237.
126. Svenningsson P, Tzavara ET, Witkin JM, et al. Involvement of striatal and extrastriatal DARPP-32 in biochemical and behavioral effects of fluoxetine (Prozac). *Proc Natl Acad Sci U S A* 2002; 99: 3182–3187.
127. Li X, Witkin JM, Need AB, Skolnick P. Enhancement of antidepressant potency by a potentiator of AMPA receptors. *Cell Mol Neurobiol* 2003; 23: 419–430.
128. Skolnick P, Legutko B, Li X, Bymaster FP. Current perspectives on the development of non-biogenic amine-based antidepressants. *Pharmacol Res* 2001; 43: 411–423.
129. Lauterborn J, Lynch G, Vanderklish P, et al. Positive modulation of AMPA receptors increases neurotrophin expression by hippocampal and cortical neurons. *J Neurosci* 2000; 20: 8–21.
130. Coyle JT, Duman RS. Finding the intracellular signaling pathways affected by mood disorder treatments. *Neuron* 2003; 38: 157–160.
131. Duman RS, Heninger GR, Nestler EJ. A molecular and cellular theory of depression. *Arch Gen Psychiatry* 1997; 54: 597–606.
132. Lauterborn J, Truong G, Baudry M, et al. Chronic elevation of brain-derived neurotrophic factor by ampakines. *J Pharmacol Exp Ther* 2003; 307: 297–305.
133. Du J, Gray NA, Falke C, et al. Structurally dissimilar antimanic agents modulate synaptic plasticity by regulating AMPA glutamate receptor subunit GluR1 synaptic expression. *Ann N Y Acad Sci* 2003; 1003: 378–380.
134. Donevan SD, Rogawski MA. GYKI 52466, a 2,3-benzodiazepine, is a highly selective, noncompetitive antagonist of AMPA/kainate receptor responses. *Neuron* 1993; 10: 51–59.
135. Rogawski MA. Molecular targets versus models for new antiepileptic drug discovery. *Epilepsy Res* 2006; 68: 22–28.
136. Nielsen EO, Varming T, Mathiesen C, et al. SPD 502: a water-soluble and in vivo long-lasting AMPA antagonist with neuroprotective activity. *J Pharmacol Exp Ther* 1999; 289: 1492–1501.
137. Conn PJ, Pin JP. Pharmacology and functions of metabotropic glutamate receptors. *Annu Rev Pharmacol Toxicol* 1997; 37: 205–237.
138. Salinska E, Stafiej A. Metabotropic glutamate receptors (mGluRs) are involved in early phase of memory formation: possible role of modulation of glutamate release. *Neurochem Int* 2003; 43: 469–474.
139. Tan Y, Hori N, Carpenter DO. The mechanism of presynaptic long-term depression mediated by group I metabotropic glutamate receptors. *Cell Mol Neurobiol* 2003; 23: 187–203.

140. Riedel G, Platt B, Micheau J. Glutamate receptor function in learning and memory. *Behav Brain Res* 2003; 140: 1–47.
141. Li X, Need AB, Baez M, Witkin JM. Metabotropic glutamate 5 receptor antagonism is associated with antidepressant-like effects in mice. *J Pharmacol Exp Ther* 2006; 319: 254–259.
142. Wieronska JM, Szewczyk B, Branski P, et al. Antidepressant-like effect of MPEP, a potent, selective and systemically active mGlu5 receptor antagonist in the olfactory bulbectomized rats. *Amino Acids* 2002; 23: 213–216.
143. Legutko B, Szewczyk B, Pomierny-Chamiolo L, et al. Effect of MPEP treatment on brain-derived neurotrophic factor gene expression. *Pharmacol Rep* 2006; 58: 427–430.
144. Belozertseva IV, Kos T, Popik P, et al. Antidepressant-like effects of mGluR1 and mGluR5 antagonists in the rat forced swim and the mouse tail suspension tests. *Eur Neuropsychopharmacol* 2007; 17: 172–179.
145. Porter RH, Jaeschke G, Spooren W, et al. Fenobam: a clinically validated nonbenzodiazepine anxiolytic is a potent, selective, and noncompetitive mGlu5 receptor antagonist with inverse agonist activity. *J Pharmacol Exp Ther* 2005; 315: 711–721.
146. Kinney GG, O'Brien JA, Lemaire W, et al. A novel selective positive allosteric modulator of metabotropic glutamate receptor subtype 5 has in vivo activity and antipsychotic-like effects in rat behavioral models. *J Pharmacol Exp Ther* 2005; 313: 199–206.
147. Chaki S, Yoshikawa R, Hirota S, et al. MGS0039: a potent and selective group II metabotropic glutamate receptor antagonist with antidepressant-like activity. *Neuropharmacology* 2004; 46: 457–467.
148. Yoshimizu T, Shimazaki T, Ito A, Chaki S. An mGluR2/3 antagonist, MGS0039, exerts antidepressant and anxiolytic effects in behavioral models in rats. *Psychopharmacology (Berl)* 2006; 186: 587–593.
149. Yoshimizu T, Chaki S. Increased cell proliferation in the adult mouse hippocampus following chronic administration of group II metabotropic glutamate receptor antagonist, MGS0039. *Biochem Biophys Res Commun* 2004; 315: 493–496.
150. Karasawa J, Shimazaki T, Kawashima N, Chaki S. AMPA receptor stimulation mediates the antidepressant-like effect of a group II metabotropic glutamate receptor antagonist. *Brain Res* 2005; 1042: 92–98.
151. Bergink V, Westenberg HG. Metabotropic glutamate II receptor agonists in panic disorder: a double blind clinical trial with LY354740. *Int Clin Psychopharmacol* 2005; 20: 291–293.
152. Danysz W. LY-544344. Eli Lilly. *IDrugs* 2005; 8: 755–762.
153. Krystal JH, Abi-Saab W, Perry E, et al. Preliminary evidence of attenuation of the disruptive effects of the NMDA glutamate receptor antagonist, ketamine, on working memory by pretreatment with the group II metabotropic glutamate receptor agonist, LY354740, in healthy human subjects. *Psychopharmacology (Berl)* 2005; 179: 303–309.
154. Gasparini F, Bruno V, Battaglia G, et al. (R,S)-4-phosphonophenylglycine, a potent and selective group III metabotropic glutamate receptor agonist, is anticonvulsive and neuroprotective in vivo. *J Pharmacol Exp Ther* 1999; 289: 1678–1687.
155. Cryan JF, Kelly PH, Neijt HC, et al. Antidepressant and anxiolytic-like effects in mice lacking the group III metabotropic glutamate receptor mGluR7. *Eur J Neurosci* 2003; 17: 2409–2417.
156. Palucha A, Tatarczynska E, Branski P, et al. Group III mGlu receptor agonists produce anxiolytic- and antidepressant-like effects after central administration in rats. *Neuropharmacology* 2004; 46: 151–159.
157. Belanoff JK, Rothschild A, Cassidy F, et al. An open label trial of C-1073 (mifepristone) for psychotic major depression. *Biol Psychiatry* 2002; 52: 386–392.
158. Arana GW, Forbes RA. Dexamethasone for the treatment of depression: a preliminary report. *J Clin Psychiatry* 1991; 52: 304–306.
159. Arana GW, Santos AB, Laraia MT, et al. Dexamethasone for the treatment of depression: a randomized, placebo-controlled, double-blind trial. *Am J Psychiatry* 1995; 152: 265–267.
160. Quiroz JA, Singh J, Gould TD, et al. Emerging experimental therapeutics for bipolar disorder: clues from the molecular pathophysiology. *Mol Psychiatry* 2004; 9: 756–776.
161. Jahn H, Schick M, Kiefer F, et al. Metyrapone as additive treatment in major depression: a double-blind and placebo-controlled trial. *Arch Gen Psychiatry* 2004; 61: 1235–1244.
162. Zobel AW, Nickel T, Kunzel HE, et al. Effects of the high-affinity corticotropin-releasing hormone receptor 1 antagonist R121919 in major depression: the first 20 patients treated. *J Psychiatr Res* 2000; 34: 171–181.
163. Schmidt PJ, Daly R, Bloch M, et al. Dehydroepiandrosterone monotherapy in midlife-onset major and minor depression. *Arch Gen Psychiatry* 2005; 62: 154–162.
164. Kunzel HE, Zobel AW, Nickel T, et al. Treatment of depression with the CRH-1-receptor antagonist R121919: endocrine changes and side effects. *J Psychiatr Res* 2003; 37: 525–533.
165. Dean CE. Prasterone (DHEA) and mania. *Ann Pharmacother* 2000; 34: 1419–1422.
166. Brown ES, Bobadilla L, Rush AJ. Ketoconazole in bipolar patients with depressive symptoms: a case series and literature review. *Bipolar Disord* 2001; 3: 23–29.
167. Carroll BJ, Rubin RT. Mifepristone in psychotic depression? *Biol Psychiatry* 2008; 63: e1.
168. Young AH, Gallagher P, Watson S, et al. Improvements in neurocognitive function and mood following adjunctive treatment with mifepristone (RU-486) in bipolar disorder. *Neuropsychopharmacology* 2004; 29: 1538–1545.
169. Rothschild AJ. Challenges in the treatment of depression with psychotic features. *Biol Psychiatry* 2003; 53: 680–690.
170. Grunberg SM, Weiss MH, Russell CA, et al. Long-term administration of mifepristone (RU486): clinical tolerance during extended treatment of meningioma. *Cancer Invest* 2006; 24: 727–733.
171. Saunders J, Williams J. Antagonists of the corticotropin releasing factor receptor. *Prog Med Chem* 2003; 41: 195–247.
172. Holmes A, Heilig M, Rupniak NM, et al. Neuropeptide system as novel therapeutic targets for depression and anxiety disorders. *Trends Pharmacol Sci* 2003; 24: 580–588.

173. Habib KE, Weld KP, Rice KC, et al. Oral administration of a corticotropin-releasing hormone receptor antagonist significantly attenuates behavioral, neuroendocrine, and autonomic responses to stress in primates. *Proc Natl Acad Sci U S A* 2000; 97: 6079–6084.
174. Ducottet C, Griebel G, Belzung C. Effects of the selective nonpeptide corticotropin-releasing factor receptor 1 antagonist antalarmin in the chronic mild stress model of depression in mice. *Prog Neuropsychopharmacol Biol Psychiatry* 2003; 27: 625–631.
175. Jutkiewicz EM, Wood SK, Houshyar H, et al. The effects of CRF antagonists, antalarmin, CP154,526, LWH234, and R121919, in the forced swim test and on swim-induced increases in adrenocorticotropin in rats. *Psychopharmacology (Berl)* 2005; 180: 215–223.
176. Seymour PA, Schmidt AW, Schulz DW. The pharmacology of CP-154,526, a non-peptide antagonist of the CRH1 receptor: a review. *CNS Drug Rev* 2003; 9: 57–96.
177. Mansbach RS, Brooks EN, Chen YL. Antidepressant-like effects of CP-154,526, a selective CRF1 receptor antagonist. *Eur J Pharmacol* 1997; 323: 21–26.
178. Griebel G, Simiand J, Steinberg R, et al. 4-(2-Chloro-4-methoxy-5-methylphenyl)-N-[(1S)-2-cyclopropyl-1-(3-fluoro-4-methylphenyl)ethyl]5-methyl-N-(2-propynyl)-1,3-thiazol-2-amine hydrochloride (SSR125543A), a potent and selective corticotrophin-releasing factor1 receptor antagonist. II. Characterization in rodent models of stress-related disorders. *J Pharmacol Exp Ther* 2002; 301: 333–345.
179. Alonso R, Griebel G, Pavone G, et al. Blockade of CRF(1) or V(1b) receptors reverses stress-induced suppression of neurogenesis in a mouse model of depression. *Mol Psychiatry* 2004; 9: 278–286.
180. Farrokhi C, Blanchard DC, Griebel G, et al. Effects of the CRF1 antagonist SSR125543A on aggressive behaviors in hamsters. *Pharmacol Biochem Behav* 2004; 77: 465–469.
181. Harro J, Tonissaar M, Eller M. The effects of CRA 1000, a non-peptide antagonist of corticotropin-releasing factor receptor type 1, on adaptive behaviour in the rat. *Neuropeptides* 2001; 35: 100–109.
182. Li Y-W, Hill G, Wong H, et al. Receptor occupancy of nonpeptide corticotropin-releasing factor 1 antagonist DMP696: correlation with drug exposure and anxiolytic efficacy. *J Pharmacol Exp Ther* 2003; 305: 86–96.
183. Devries AC, Joh HD, Bernard O, et al. Social stress exacerbates stroke outcome by suppressing Bcl-2 expression. *Proc Natl Acad Sci U S A* 2001; 98: 11824–11828.
184. Takata K, Kitamura Y, Kakimura J, et al. Increase of bcl-2 protein in neuronal dendritic processes of cerebral cortex and hippocampus by the antiparkinsonian drugs, talipexole and pramipexole. *Brain Res* 2000; 872: 236–241.
185. Goldberg JF, Burdick KE, Endick CJ. Preliminary randomized, double-blind, placebo-controlled trial of pramipexole added to mood stabilizers for treatment-resistant bipolar depression. *Am J Psychiatry* 2004; 161: 564–566.
186. Zarate CA Jr, Payne JL, Singh J, et al. Pramipexole for bipolar II depression: a placebo-controlled proof of concept study. *Biol Psychiatry* 2004; 56: 54–60.
187. Gould TD, Manji HK. Glycogen synthase kinase-3: a putative molecular target for lithium mimetic drugs. *Neuropsychopharmacology* 2005; 30: 1223–1237.
188. Gould TD, Einat H, Bhat R, Manji HK. AR-A014418, a selective GSK-3 inhibitor, produces antidepressant-like effects in the forced swim test. *Int J Neuropsychopharmacol* 2004; 7: 387–390.
189. Gould TD, Picchini AM, Einat H, Manji HK. Targeting glycogen synthase kinase-3 in the CNS: implications for the development of new treatments for mood disorders. *Curr Drug Targets* 2006; 7: 1399–1409.
190. Young LT, Wang JF, Woods CM, Robb JC. Platelet protein kinase C alpha levels in drug-free and lithium-treated subjects with bipolar disorder. *Neuropsychobiology* 1999; 40: 63–66.
191. Manji HK, Lenox RH. Ziskind-Somerfeld Research Award. Protein kinase C signaling in the brain: molecular transduction of mood stabilization in the treatment of manic-depressive illness. *Biol Psychiatry* 1999; 46: 1328–1351.
192. Chen G, Manji HK, Hawver DB, et al. Chronic sodium valproate selectively decreases protein kinase C alpha and epsilon in vitro. *J Neurochem* 1994; 63: 2361–2364.
193. Friedman E, Hoau Yan W, Levinson D, et al. Altered platelet protein kinase C activity in bipolar affective disorder, manic episode. *Biol Psychiatry* 1993; 33: 520–525.
194. Manji HK, Etcheberrigaray R, Chen G, Olds JL. Lithium decreases membrane-associated protein kinase C in hippocampus: selectivity for the alpha isozyme. *J Neurochem* 1993; 61: 2303–2310.
195. Hahn CG, Friedman E. Abnormalities in protein kinase C signaling and the pathophysiology of bipolar disorder. *Bipolar Disord* 1999; 1: 81–86.
196. Hilakivi-Clarke L. Role of estradiol in alcohol intake and alcohol-related behaviors. *J Stud Alcohol* 1996; 57: 162–170.
197. Einat H, Yuan P, Szabo ST, et al. Protein kinase C inhibition by tamoxifen antagonizes manic-like behavior in rats: implications for the development of novel therapeutics for bipolar disorder. *Neuropsychobiology* 2007; 55: 123–131.
198. Bebchuk JM, Arfken CL, Dolan-Manji S, et al. A preliminary investigation of a protein kinase C inhibitor in the treatment of acute mania. *Arch Gen Psychiatry* 2000; 57: 95–97.
199. Zarate CA Jr, Singh JB, Carlson PJ, et al. Efficacy of a protein kinase C inhibitor (tamoxifen) in the treatment of acute mania: a pilot study. *Bipolar Disord* 2007; 9: 561–570.
200. Kulkarni J, Garland KA, Scaffidi A, et al. A pilot study of hormone modulation as a new treatment for mania in women with bipolar affective disorder. *Psychoneuroendocrinology* 2006; 31: 543–547.
201. Goldstein JA. Danazol and the rapid-cycling patient. *J Clin Psychiatry* 1986; 47: 153–154.
202. Rapoport SI, Bosetti F. Do lithium and anticonvulsants target the brain arachidonic acid cascade in bipolar disorder? *Arch Gen Psychiatry* 2002; 59: 592–596.
203. Bosetti F, Weerasinghe GR, Rosenberger TA, Rapoport SI. Valproic acid down-regulates the conversion of arachidonic acid to eicosanoids via cyclooxygenase-1 and -2 in rat brain. *J Neurochem* 2003; 85: 690–696.

204. Rao JS, Bazinet RP, Rapoport SI, Lee HJ. Chronic treatment of rats with sodium valproate downregulates frontal cortex NF-kappaB DNA binding activity and COX-2 mRNA. *Bipolar Disord* 2007; 9: 513–520.
205. Myint AM, Steinbusch HW, Goeghegan L, et al. Effect of the COX-2 inhibitor celecoxib on behavioural and immune changes in an olfactory bulbectomised rat model of depression. *Neuroimmunomodulation* 2007; 14: 65–71.
206. Muller N, Schwarz MJ, Dehning S, et al. The cyclooxygenase-2 inhibitor celecoxib has therapeutic effects in major depression: results of a double-blind, randomized, placebo controlled, add-on pilot study to reboxetine. *Mol Psychiatry* 2006; 11: 680–684.
207. Nery FG, Monkul ES, Hatch JP, et al. Celecoxib as an adjunct in the treatment of depressive or mixed episodes of bipolar disorder: a double-blind, randomized, placebo-controlled study. *Hum Psychopharmacol* 2008; 23: 87–94.
208. Velentgas P, West W, Cannuscio CC, et al. Cardiovascular risk of selective cyclooxygenase-2 inhibitors and other non-aspirin non-steroidal anti-inflammatory medications. *Pharmacoepidemiol Drug Saf* 2006; 15: 641–652.
209. Littner M, Johnson SF, McCall WV, et al. Practice parameters for the treatment of narcolepsy: an update for 2000. *Sleep* 2001; 24: 451–466.
210. Ferraro L, Tanganelli S, O'Connor WT, et al. The vigilance promoting drug modafinil increases dopamine release in the rat nucleus accumbens via the involvement of a local GABAergic mechanism. *Eur J Pharmacol* 1996; 306: 33–39.
211. Frye MA, Grunze H, Suppes T, et al. A placebo-controlled evaluation of adjunctive modafinil in the treatment of bipolar depression. *Am J Psychiatry* 2007; 164: 1242–1249.
212. Post RM, Altshuler LL, Frye MA, et al. New findings from the Bipolar Collaborative Network: clinical implications for therapeutics. *Curr Psychiatry Rep* 2006; 8: 489–497.
213. Wolf J, Fiedler U, Anghelescu I, Schwertfeger N. Manic switch in a patient with treatment-resistant bipolar depression treated with modafinil. *J Clin Psychiatry* 2006; 67: 1817.
214. Ranjan S, Chandra PS. Modafinil-induced irritability and aggression? A report of 2 bipolar patients. *J Clin Psychopharmacol* 2005; 25: 628–629.
215. Goodnick PJ. Anticonvulsants in the treatment of bipolar mania. *Expert Opin Pharmacother* 2006; 7: 401–410.
216. Zarate CA, Manji H. Protein kinase C inhibitors: rationale for use and potential in the treatment of bipolar disorder. *CNS Drugs* 2008; Available at: http://cnsdrugs.adisonline.com.

Anxiety Disorders: A Comprehensive Review of Pharmacotherapies

Ellen J. Hoffman, MD,[1] and Sanjay J. Mathew, MD[2]

[1]Division of Child and Adolescent Psychiatry, Department of Psychiatry, Mount Sinai School of Medicine, New York, NY
[2]Mood and Anxiety Disorders Program, Department of Psychiatry, Mount Sinai School of Medicine, New York, NY

ABSTRACT

This article reviews the evidence from randomized, placebo-controlled trials and meta-analyses of pharmacological treatments of the following anxiety disorders: generalized anxiety disorder, panic disorder, social anxiety disorder, and post–traumatic stress disorder. There is evidence from multiple randomized, placebo-controlled trials to support the use of selective serotonin reuptake inhibitors as first-line pharmacotherapy in these disorders, and a number of the selective serotonin reuptake inhibitors have received US Food and Drug Administration approval for these indications. Serotonin-norepinephrine reuptake inhibitors are now emerging as first-line treatments for these anxiety disorders alongside the selective serotonin reuptake inhibitors and have been US Food and Drug Administration–approved for some of these indications as well. Benzodiazepines are also effective treatments for anxiety disorders, and although this medication class has the advantage of a rapid onset of action, their use is limited by their potential for abuse and lack of antidepressant properties. In addition to reviewing the clinical trials that have investigated the anxiolytic effects of these commonly used medications, we review the evidence for novel uses of other agents, including anticonvulsants and atypical antipsychotics, in anxiety disorders. ***Mt Sinai J Med 75:248–262, 2008.*** © *2008 Mount Sinai School of Medicine*

Address Correspondence to:

Sanjay J. Mathew
Department of Psychiatry
Mount Sinai School of Medicine
One Gustave L. Levy Place Box 1217
New York, NY 10019
Email: sanjay.mathew@mssm.edu

Key Words: benzodiazepines, generalized anxiety disorder, panic disorder, post–traumatic stress disorder, pharmacotherapy, selective norepinephrine reuptake inhibitors, selective serotonin reuptake inhibitors, social anxiety disorder.

The anxiety disorders with the greatest evidence for the efficacy of pharmacotherapy are generalized anxiety disorder (GAD), panic disorder (PD), social anxiety disorder (SAD), and obsessive-compulsive disorder, whereas pharmacotherapy evidence for post–traumatic stress disorder (PTSD) is more limited. The therapeutics of obsessive-compulsive disorder are discussed elsewhere in this issue (see Berlin and Hollander), whereas the treatment of specific phobias has been excluded because of space considerations. The medication classes most commonly used for GAD, PD, SAD, and PTSD are selective serotonin reuptake inhibitors (SSRIs) and serotonin-norepinephrine reuptake inhibitors (SNRIs), whereas other pharmacotherapy approaches include benzodiazepines (BZDs), tricyclic antidepressants (TCAs), monoamine oxidase inhibitors (MAOIs), anticonvulsants, and atypical antipsychotics.

SSRIs are considered first-line medications for the treatment of GAD, PD, SAD, and PTSD.[1–4] Although older classes of antidepressant medications–TCAs and MAOIs–have been shown to be effective for anxiety disorders, SSRIs are the most commonly prescribed because of their more favorable side-effect profile. In addition, unlike BZDs, SSRIs have the advantage of treating comorbid depression, whereas long-term administration of SSRIs is not associated with the potential for abuse or physiological dependence. Table 1 shows the non-BZD medications that have been approved by the US Food and Drug Administration (FDA) for the treatment of the anxiety disorders. Table 2 lists the indications and pharmacological properties of commonly used BZDs.

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20041

© 2008 Mount Sinai School of Medicine

Table 1. *US Food and Drug Administration–Approved Non-Benzodiazepine Medications for Anxiety Disorders in Adults.*

Medication	GAD	PD	PTSD	SAD	Daily Dose Range	Other Approved Psychiatric Indications
SSRIs						
Escitalopram (Lexapro)	✓	–	–	–	10–20 mg	MDD
Fluoxetine (Prozac)	–	✓	–	–	20–60 mg	MDD, OCD, bulimia nervosa, PMDD (Sarafem)
Paroxetine (Paxil)	✓	✓	✓	✓	PD: 10–60 mg SAD: 20–60 mg GAD, PTSD: 20–50 mg	MDD, OCD
Paroxetine CR (Paxil CR)	–	✓	–	✓	PD: 12.5–75 mg SAD: 12.5–37.5 mg	MDD, PMDD
Sertraline (Zoloft)	–	✓	✓	✓	50–200 mg	MDD, OCD, PMDD
SNRIs						
Duloxetine (Cymbalta)	✓	–	–	–	30–120 mg	MDD
Venlafaxine XR (Effexor XR)	✓	✓	–	✓	37.5–225 mg	MDD
Azapirone						
Buspirone (BuSpar)	✓*	–	–	–	15–60 mg	Short-term relief of anxiety symptoms

NOTE: This table includes anxiety disorders, excluding OCD, from the *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed., text revision, and was adapted from Mathew *et al.*[103]

Abbreviations: GAD, generalized anxiety disorder; MDD, major depressive disorder; OCD, obsessive-compulsive disorder; PD, panic disorder; PMDD, premenstrual dysphoric disorder; PTSD, post–traumatic stress disorder; SAD, social anxiety disorder; SNRI, serotonin-norepinephrine reuptake inhibitor; SSRI, selective serotonin reuptake inhibitor.

*Approved for anxiety disorders that correspond most closely to GAD as described in the *Diagnostic and Statistical Manual of Mental Disorders*, 3rd ed., or for anxiety disorders with no individual disorder specified.

GENERALIZED ANXIETY DISORDER

The diagnostic criteria for GAD in the *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed., text revision (DSM-IV-TR), include excessive anxiety and worry, difficulty controlling the worry, and symptoms of restlessness, fatigability, poor concentration, irritability, muscle tension, and poor sleep on most days for a period of 6 months.[5] In clinical trials, the therapeutic response to pharmacotherapy in GAD is assessed with the Hamilton Rating Scale for Anxiety (HAM-A), a 14-question scale in which patients rate the severity of their symptoms, including anxious mood, tension, insomnia, and depressed mood, along with somatic symptoms, such as cardiovascular (palpitations and chest pain), gastrointestinal (nausea or vomiting), and respiratory symptoms (shortness of breath).[6] The Clinical Global Impression Improvement (CGI-I) and Clinical Global Impression Severity (CGI-S) scales are clinician-rated scales that are also used in assessing the response in GAD as well as other psychiatric disorders.[7]

SSRIs and SNRIs

Multiple double-blind, placebo-controlled randomized clinical trials (RCTs), using large multicenter sample sizes, have demonstrated the efficacy of several SSRIs in GAD (reviewed by Mathew and Hoffman[1]). Paroxetine and escitalopram are the SSRIs that are approved by the FDA for the treatment of GAD. Two 8-week RCTs found that paroxetine, in fixed (20 or 40 mg/day) or flexible (20–50 mg/day) doses, led to significant reductions in HAM-A versus placebo (8 and 9, respectively). (Please note that in this review, when we state that study results are "significant," we are referring to statistical significance: $P < 0.05$.) Both studies found that a significant percentage of patients treated with paroxetine (62%–68%) received CGI-I scores indicating that they were "much improved" or "very much improved" after 2 months in comparison with approximately 47% of patients in the placebo group.[8,9] Also, a significant number of patients (approximately one-third) who were treated with paroxetine attained remission (defined as HAM-A ≤ 7) by the end of these studies.[8,9] Escitalopram (10–20 mg/day) was similarly found to be effective in GAD in a flexibly dosed 8-week RCT, leading to significant reductions in HAM-A scores beginning in the first week of treatment as well as significant CGI-I response rates versus placebo (58% versus 38%) by study end.[10] Davidson *et al.*[11] investigated the long-term efficacy of escitalopram in GAD in a 6-month open-label extension study for completers of three 8-week, double-blind, controlled

Table 2. *Characteristics of Commonly Used Benzodiazepines.*

Generic Name (Brand Name)	FDA-Approved Psychiatric Indications	Routes of Administration	Oral Dose Equivalency (mg)	Approved Oral Dose Range (mg)	Rate of Onset After Oral Dose	Elimination Half-Life (Hours)	Strengths (mg) and Available Preparations
Alprazolam (Xanax)	GAD;[*] PD; short-term relief of anxiety symptoms; anxiety associated with depression Alprazolam XR (Xanax XR): PD	PO	0.5	0.75–4 Alprazolam XR: 3–6 (suggested), 1–10 (used in clinical trials)	Intermediate	6–20	0.25, 0.5, 1, 2 Alprazolam Intensol (concentrated solution): 1 mg/mL Alprazolam XR: 0.5, 1, 2, 3 Niravam (alprazolam orally disintegrating): 0.25, 0.5, 1, 2
Chlordiazepoxide (Librium)	Anxiety disorders;[†] short-term relief of anxiety symptoms; ETOH withdrawal; preoperative anxiety	PO	10.0	15–100	Intermediate	30–100	5, 10, 25
Clonazepam (Klonopin)	PD	PO	0.25	0.5–4	Intermediate	18–40	0.5, 1, 2 Klonopin wafers: 0.125, 0.25, 0.5, 1, 2
Clorazepate (Tranxene)	Anxiety disorders;[†] short-term relief of anxiety symptoms; ETOH withdrawal	PO	7.5	15–60 Tranxene SD: 11.25–45	Rapid	30–100	Tranxene T-tab: 3.75, 7.5, 15 Tranxene SD: 11.25, 22.5
Diazepam (Valium)	Anxiety disorders;[†] short-term relief of anxiety symptoms; ETOH withdrawal	PO, PR, IM, IV	5.0	4–40	Rapid	30–100	2, 5, 10 Solution: 5 mg/5 mL Intensol: 5 mg/mL Diastat (diazepam rectal): 2.5 gel AcuDial gel: 10, 20 IM, IV
Lorazepam (Ativan)	Anxiety disorders;[†] short-term relief of anxiety symptoms; anxiety associated with depression	PO, IM, IV	1.0	1–10	Intermediate	10–20	0.5, 1, 2 Concentrated: 2 mg/mL IM, IV
Oxazepam (Serax)	Anxiety disorders;[†] short-term relief of anxiety symptoms; anxiety associated with depression; anxiety in older patients; ETOH withdrawal	PO	15.0	30–120	Intermediate-slow	8–12	10, 15, 30
Temazepam (Restoril)	Insomnia, short-term	PO	30.0	7.5–30	Intermediate	8–20	7.5, 15, 30
Triazolam (Halcion)	Insomnia, short-term	PO	0.25	0.125–0.5	Intermediate	2–5	0.125, 0.25

NOTE: This table was adapted from Mathew *et al.*,[103] Rosenbaum *et al.*,[129] and Goddard *et al.*[130]

Abbreviations: ETOH, ethanol; FDA, US Food and Drug Administration; GAD, generalized anxiety disorder; IM, intramuscular; IV, intravenous; PO, per os; PD, panic disorder; PR, per rectum; SD, single dose.

[*] Approved for anxiety disorders that correspond most closely to GAD as described in the *Diagnostic and Statistical Manual of Mental Disorders*, 3rd ed.

[†] No individual disorder is specified.

escitalopram trials and showed ongoing improvement in HAM-A and quality-of-life scores in the open-label phase, in which the last visit of the lead-in trial was defined as baseline, such that 76% of patients in the extension trial were responders and 49% were remitters.

Relapse prevention studies demonstrated that in comparison with patients who continued to receive escitalopram or paroxetine for 6 months, the risk of relapse was approximately 4 to 5 times greater in the placebo groups.[12,13] There are also studies demonstrating the efficacy of sertraline in GAD,[14,15] although it has not been approved by the FDA for this indication. With respect to other SSRIs (citalopram, fluoxetine, and fluvoxamine), although there is evidence that these medications treat anxiety symptoms in depressed patients, to our knowledge, there are no published RCTs of these medications for GAD.

Along with SSRIs, SNRIs are emerging as first-line medications in treating anxiety disorders. Both venlafaxine XR and recently duloxetine (in 2007) have gained FDA approval for the treatment of GAD. There are short- and long-term RCTs of venlafaxine XR in GAD. Two 6-month RCTs of fixed (37.5, 75, and 150 mg/day) or flexible (75–225 mg/day) doses of venlafaxine XR found significant reductions in HAM-A by study end in the medication groups versus the placebo groups.[16,17] Allgulander *et al.*[16] showed that venlafaxine XR at 75 or 150 mg/day sustained significant HAM-A response rates (≥50% reduction in the total score) from week 2 onward. Gelenberg *et al.*[17] reported that venlafaxine XR (75–225 mg/day) led to significant HAM-A response rates (≥40% reduction in the total score) beginning at week 1. In the short-term treatment of GAD, Rickels *et al.*[18] conducted an 8-week RCT of venlafaxine XR (75, 150, or 225 mg/day) and concluded that only the group receiving the 225 mg/day dose saw a significant reduction in total HAM-A scores by study end. In another 8-week, placebo-controlled RCT, Davidson *et al.*[19] compared venlafaxine XR (75 or 150 mg/day) and the azapirone buspirone (30 mg/day, dosed as 10 mg three times a day) and found that CGI-I response rates, but not HAM-A response rates, were significantly greater only in the 75 mg/day venlafaxine XR and buspirone groups in comparison with the placebo group by study end. However, both venlafaxine XR doses were significantly better than buspirone and placebo on the Hospital Anxiety and Depression Anxiety subscale, which is patient-rated, by week 3.[19] A relatively small, 8-week RCT with 46 patients with GAD described significant response (defined as ≥50% reduction in HAM-A; 92%) and remission (63%) rates in the venlafaxine XR (75–150 mg/day) group versus the placebo group.[20] Another 8-week RCT, however, found no significant differences in HAM-A or CGI-I scores between placebo and venlafaxine XR (75 or 150 mg/day) or diazepam (15 mg/day); a secondary analysis that omitted study centers in which there was no difference between diazepam and placebo showed significant HAM-A and CGI-I reductions in both venlafaxine XR groups versus placebo.[21]

Duloxetine was found to be effective in GAD in two 10-week, multicenter RCTs.[22,23] Rynn *et al.*[22] found that duloxetine (60–120 mg/day, progressively titrated) led to a significant decrease in HAM-A, although no significant differences in remission rates were observed. In addition, there were significant increases in the heart rate, blood pressure, and some liver measures in the duloxetine group in comparison with placebo, but the magnitude of the changes was not thought to be clinically relevant.[22] In a placebo-controlled, head-to-head RCT comparing duloxetine (60–120 mg/day) and venlafaxine XR (75–225 mg/day), there was a significant decrease in total HAM-A scores in both active treatment groups beginning at weeks 1 and 2, respectively.[23] However, Hartford *et al.*[23] noted that in comparison with placebo, the percentage of patients achieving remission was greater in the venlafaxine XR group ($P \geq 0.05$) but not in the duloxetine group.

Additional head-to-head studies of SSRIs and SNRIs in GAD are limited by the absence of placebo controls. One 6-month study showed that both escitalopram (10–20 mg/day) and paroxetine (20–50 mg/day) led to similar HAM-A and quality-of-life improvements, although escitalopram was better tolerated than paroxetine.[24] A 2-month head-to-head study found that sertraline (25–100 mg/day) and paroxetine (10–40 mg/day) led to similar response (61%–68%) and remission (40%–46%) rates.[25] In a small, 8-week, open-label study, both venlafaxine XR (37.5–225 mg/day) and paroxetine (10–40 mg/day) led to significant HAM-A and CGI-S reductions over time.[26] Of note, paroxetine led to significantly more weight gain than venlafaxine XR, whereas venlafaxine XR resulted in significant increases in the mean change of systolic and diastolic blood pressure from the baseline to the endpoint in comparison with paroxetine.[26]

BZDs and Buspirone

Multiple RCTs have demonstrated the short- and long-term efficacy of BZDs in anxiety disorders, including GAD, PD, and SAD (reviewed by Mathew and Hoffman,[1] Katon,[2] and Davidson[3]). BZDs are effective in the acute treatment of GAD and have a rapid onset of action.[27,28] However, long-term use of BZDs in GAD is problematic because few patients achieve and sustain remission with BZD monotherapy (reviewed by Gorman[29]). Advantages of BZDs over SSRIs and SNRIs include their ability to be used on an "as-needed" basis for anxiety and their rapid onset of action, which leads to their frequent use in clinical practice as adjunctive medications for stabilization during initiation of an SSRI or SNRI. In addition, longer acting BZDs, such as clonazepam or alprazolam XR, may be used to decrease breakthrough anxiety (reviewed by Katon[2]). Studies have also shown that BZDs preferentially affect the somatic symptoms in GAD,[27,28] whereas SSRIs have been found to have a greater effect on the psychic symptoms.[8,9,25] However, the disadvantages of BZDs are the risks of abuse and physiological dependence with long-term use,[30] as well as their lack of antidepressant effects, which is important given that depression and GAD are often comorbid.[29]

Buspirone is an azapirone, a serotonin 1A (5-hydroxytryptamine receptor 1A) agonist, approved by the FDA for treating "anxiety disorders," a broad diagnosis corresponding most closely to GAD in the *Diagnostic and Statistical Manual of Mental Disorders*, 3rd ed. (DSM-III). A retrospective meta-analysis of pooled data from 8 placebo-controlled studies found that buspirone resulted in a significant improvement in HAM-A for GAD.[31] Advantages of buspirone include its overall tolerability and lack of addictive potential (reviewed by Mathew and Hoffman[1]), although an important limitation is evidence of decreased efficacy in patients with recent prior BZD use.[32] Also, RCTs have shown that buspirone is not effective in PD[33–35] or SAD.[36]

Anticonvulsants

There is emerging evidence from RCTs for pregabalin in GAD (reviewed by Mula *et al.*[37]). Although its exact mechanism of action is unknown, pregabalin binds to the $\alpha_2\delta$ subunit of voltage-gated calcium channels, which reduces neurotransmitter release. Pregabalin has been approved by the FDA as an adjunctive treatment for partial seizures as well as diabetic peripheral neuropathy, postherpetic neuralgia, and fibromyalgia; in 2006, it was approved by the European Commission for treating GAD, but at this time, it is not approved by the FDA for this indication.

RCTs have demonstrated the efficacy of pregabalin in GAD after 4 and 6 weeks, although long-term studies are still needed. A recent dose-response analysis by Bech[38] pooled data from 4 short-term, fixed-dose RCTs of pregabalin in GAD[39–42] and concluded that pregabalin in the range of 200 to 450 mg/day resulted in HAM-A effect sizes (ESs) > 0.40 with a "plateau-like response curve." However, a lower dose of pregabalin (150 mg/day) led to an ES < 0.40, and at a higher dose (600 mg/day), ESs were similar to those found for 200 to 450 mg/day pregabalin.[38] One potential advantage of pregabalin is that it appears to have a relatively rapid onset of action. In a 6-week RCT, all 3 pregabalin doses tested (200 mg/day divided into 2 doses a day, 400 mg/day divided into 2 doses a day, and 450 mg/day divided into 3 doses a day) resulted in significant HAM-A reductions in comparison with placebo beginning at week 1.[41] Another 6-week RCT in GAD, which compared pregabalin doses of 400 or 600 mg/day divided into 2 doses a day and a venlafaxine dose of 75 mg/day divided into 2 doses a day, reported significant improvements in HAM-A scores versus placebo beginning at week 1 in both pregabalin groups and at week 2 in the venlafaxine group.[43]

Additionally, in 4-week RCTs, pregabalin appears to demonstrate efficacy comparable to that of BZDs in reducing HAM-A,[39,40,42] while apparently resulting in less withdrawal. During discontinuation, Pande *et al.*[40] reported no significant difference in Physician Withdrawal Checklist scores for pregabalin versus placebo, unlike lorazepam; other studies reported some significant increases in Physician Withdrawal Checklist scores for pregabalin versus placebo, although these changes were not considered clinically significant and appeared to be dose-related.[39,41,42] The most commonly reported side effects of pregabalin include somnolence and dizziness and weight gain in patients taking higher doses.

Evidence for the anticonvulsant tiagabine (Gabatril), a gamma-aminobutyric acid reuptake inhibitor, in GAD is very limited (reviewed by Mathew and Hoffman[1]).

Atypical Antipsychotics

There are a few RCTs that have investigated the role of atypical antipsychotics in anxiety disorders, as these medications have been considered adjunctive treatments in these disorders because of their broad neurochemical effects on postsynaptic 5-hydroxytryptamine 2 receptors and the modulation of 5-hydroxytryptamine receptor 1A (reviewed by Mathew and Hoffman[1]). Two very small RCTs investigated olanzapine and risperidone as adjunctive agents in patients with GAD who did not respond to another medication, either fluoxetine in the olanzapine study[44] or another medication (an SSRI, SNRI, BZD, or other anxiolytic or antidepressant) in the risperidone study.[45] Although olanzapine augmentation resulted in a significant number of HAM-A responders after 6 weeks[44] and adjunctive risperidone led to significant HAM-A reductions after 5 weeks,[45] there were no significant findings for other outcome measures. In addition, olanzapine resulted in sedation and significant weight gain (11.0 lb on average),[44] and side effects of risperidone included somnolence, dizziness, and blurred vision.[45]

GAD Meta-Analysis

In a meta-analysis of RCTs of antidepressants, including imipramine, venlafaxine, and paroxetine, in adults with GAD alone, Kapczinski *et al.*[46] reported that antidepressants were more effective than placebo and that dropout rates were not different among the antidepressants. The number needed to treat for these antidepressants as a group in GAD was calculated to be between 5 and 6.[46]

PANIC DISORDER

DSM-IV-TR criteria for PD include recurrent panic attacks (periods of intense fear that peak within 10 minutes and are associated with symptoms such as palpitations, sweating, shortness of breath, chest pain, and lightheadedness) as well as persistent worry about the attacks or fear of having more attacks, which may result in a change in behavior (reviewed by Katon[2]). PD may occur with or without agoraphobia, which is described in DSM-IV-TR as anxiety about and avoidance of being in situations where escape may be difficult.[5] Clinical trials assessing the efficacy of pharmacotherapy in PD examine outcome measures such as panic attack frequency, the percentage of patients who are free of attacks, and the change in the total number of attacks. The Panic Disorder Severity Scale (PDSS), which measures the severity of PD symptoms, including panic attack frequency, distress during attacks, anticipatory anxiety, agoraphobia, avoidance of bodily sensations, and impairment of work and social functioning, is also used in some studies.[47] Of note, PD is one of the anxiety disorders that tends to have higher placebo responsivity.[48]

SSRIs and SNRIs

Both SSRIs [paroxetine, controlled-release paroxetine (paroxetine CR), sertraline, and fluoxetine] and the SNRI venlafaxine XR have been approved by the FDA to treat PD. The SSRIs are recommended as first-line medications in PD on the basis of their safety and overall tolerability, although SSRIs, TCAs, and BZDs have been shown to be equally effective in the treatment of PD and associated anxiety symptoms (reviewed by Heuer *et al.*[19]). If a patient with PD does not respond to treatment with an SSRI, a trial with another SSRI should be attempted, and if that fails, switching to a medication in a different class, such as venlafaxine XR (SNRI) or a TCA or BZD, is recommended.[19] There is also evidence for initial augmentation of SSRIs with BZDs early in the treatment of PD (see the subsection on BZDs) due to the rapid onset of action of BZDs. SSRIs should be initiated at low doses with slow titration to minimize side effects, which may include nausea, anxiety, tremors, jitteriness, sexual dysfunction, and insomnia, as patients with PD may be sensitive to side effects.[2,49]

RCTs have provided evidence for the efficacy of many SSRIs in PD. In 12- and 36-week RCTs using the TCA clomipramine as an active comparator, paroxetine was found to be effective in PD.[50,51] Also, paroxetine (40 mg/day) resulted in significant reductions in the total number of panic attacks by week 4 of a 10-week RCT.[52] A pooled analysis of three 10-week RCTs showed that paroxetine CR led to a significantly greater percentage of panic-free patients versus placebo (63% versus 53%) in the 2 weeks prior to the study endpoint.[53] RCTs have also investigated sertraline in PD.[54–56] One multicenter, 10-week RCT of sertraline (50–200 mg/day) found that sertraline led to significant reductions in the frequency of panic attacks and in PDSS scores by the study endpoint,[54] whereas another 10-week RCT reported a significant decrease in the number of panic attacks per week in the observed case analysis by week 3, and a significant percentage of patients were panic-free by the study endpoint in the sertraline group.[55] A 28-week discontinuation RCT showed that significantly fewer patients who continued on sertraline after 1 year of open-label treatment experienced an exacerbation of PD symptoms in comparison with those switched to placebo.[56]

There is also evidence from RCTs for fluoxetine,[57] fluvoxamine,[58–60] escitalopram,[61] and citalopram[62,63] in PD. In a 12-week RCT of fluoxetine (20 mg/day, increased to a maximum of 60 mg/day after 6 weeks in patients who did not respond sufficiently), Michelson *et al.*[57] reported a significant reduction in PDSS scores and a significant increase in the proportion of patients who were free of panic attacks after both 6 and 12 weeks of treatment, although there were variable results in an older fluoxetine trial in which lower doses were used.[64] Stahl *et al.*[61] compared escitalopram (10–20 mg/day) and citalopram (20–40 mg/day) in a 10-week, placebo-controlled trial and reported that escitalopram, but not citalopram, led to a significant decrease in panic attack frequency by the study endpoint, although the proportion of patients who were panic-free was significant only at the trend level for the escitalopram group ($P = 0.051$). Both escitalopram and citalopram groups demonstrated significant improvements on CGI-I after 10 weeks.[61] In a 12-week, head-to-head study of sertraline and paroxetine, which lacked a placebo arm, both medications demonstrated efficacy in PD, although there were numerically fewer dropouts due to adverse events and significantly less weight gain in the sertraline group.[65]

Venlafaxine XR (75 or 150 mg/day) and paroxetine (40 mg/day) were compared in a 12-week, placebo-controlled, double-blind trial, which found no significant differences in efficacy between the venlafaxine XR and paroxetine groups; both groups saw a significant percentage of panic-free patients by study end.[66] In another multicenter trial, venlafaxine XR (75–225 mg/day) resulted in a significant reduction in the frequency of full-symptom panic attacks in comparison with placebo after 10 weeks, and although venlafaxine XR did not result in a significantly greater number of patients who were free of panic attacks by the study endpoint, there was a significant percentage of CGI-I responders and remitters in the venlafaxine XR group.[67] Of note, there was a significant increase in heart rate in the venlafaxine XR group (change from the baseline in the supine pulse rate = 2.18 beats per minute).[67] Venlafaxine XR was also shown to be effective in preventing relapse in PD.[68]

BZDs

Although there is evidence from earlier studies supporting the use of BZD monotherapy in PD, more recent studies have examined and provided evidence for the common clinical usage of BZDs in combination with SSRIs in the acute treatment of PD.[69,70] Goddard *et al.*[69] reported that there were significantly more responders after 1 week in patients receiving sertraline plus clonazepam versus sertraline plus placebo, although there were no group differences in response during weeks 4 to 12 (clonazepam was tapered and discontinued after 3 weeks). Pollack *et al.*[70] reported a similar finding of an early benefit of combination treatment (paroxetine plus clonazepam) versus paroxetine monotherapy generally in the first 5 weeks. However, by the conclusion of the 12-week study, there was no significant difference in outcome between the monotherapy or combination groups or a group treated with combination therapy for 4 weeks followed by clonazepam taper, with a significant improvement on PDSS in all groups.

PD Meta-Analyses

Meta-analyses determined moderate ESs (~0.4–0.55) and tolerability of SSRIs comparable to those of TCAs and BZDs in PD.[71,72] Studies have also investigated the efficacy of pharmacotherapy versus psychotherapy in PD. Mitte[71] reported that cognitive-behavioral and behavioral therapies were at least as effective as pharmacotherapy, and pharmacotherapy was superior to placebo. In a Cochrane Collaboration meta-analysis of 23 randomized studies comparing psychotherapy and antidepressants in PD, Furukawa *et al.*[73] found that in the first few months of treatment, the combination of psychotherapy and pharmacotherapy was more effective than either antidepressant treatment or psychotherapy alone, and this trend persisted as long as the medication was continued. They concluded that either combination therapy or psychotherapy alone may be used as first-line treatment in PD.[73]

SOCIAL ANXIETY DISORDER

In DSM-IV-TR, generalized SAD, on which we concentrate in this article, is characterized by a persistent fear of social situations, such that exposure to these situations leads to anxiety and avoidance of these situations, which interfere with the person's functioning (reviewed by Schneier[74]). Nongeneralized or "performance-type" SAD involves fear of public speaking or performance situations only.[74] Most RCTs use the Liebowitz Social Anxiety Scale (LSAS), which assesses a patient's level of fear and avoidance of 24 distinct social situations, such as attending a party, meeting strangers, or being the center of attention, in order to study the efficacy of pharmacotherapy in treating generalized SAD.[75]

SSRIs and SNRIs

SSRIs and SNRIs are considered first-line medications for generalized SAD.[74] Paroxetine, paroxetine CR, sertraline, and venlafaxine XR are FDA-approved for treating SAD. These medications are often started at half of the usual effective dose, and the dose is increased after the first week of treatment.[74] Although their dose-response curve is relatively flat, doses of these medications are typically increased as tolerated in patients who do not respond after 4 weeks, as some patients benefit from increased doses.[74] Although many patients improve within the first few weeks of treatment, it has been suggested that initial SSRI trials should last 12 weeks, as at least 25% of those who do not respond after 8 weeks may respond during the subsequent 4 weeks at the same medication dose; to decrease the risk of relapse, maintenance pharmacotherapy for those patients who respond after 12 weeks is recommended.[74]

There are multiple RCTs supporting the efficacy of these medications in SAD. Paroxetine and paroxetine CR were shown to be effective in flexibly dosed, 12-week, multicenter RCTs of SAD.[76–78] Paroxetine (20–50 mg/day) led to significant reductions in LSAS by week 2[76] and week 4[77] in 2 studies. In a 12-week, fixed-dose RCT of paroxetine (20, 40, and 60 mg/day), only the 20 mg/day dose resulted in significant LSAS reductions by study end, although the number of CGI-I responders was significant for the 40-mg group.[79] In the study by Lepola *et al.*,[78] there was a significant reduction in LSAS by week 6 in the paroxetine CR group, and significantly more patients in this group achieved remission on CGI-I. RCTs have also demonstrated the efficacy of sertraline (50–200 mg/day) in SAD after 12[80] and 20 weeks.[81] In a multinational RCT, escitalopram (10–20 mg) led to significant reductions in LSAS after 12 weeks.[82] There is also evidence from 10- to 12-week RCTs for fluvoxamine[83,84] and fluvoxamine CR[85,86] in SAD, although a 12-week extension of a 12-week RCT showed continued but nonsignificant LSAS improvement in the fluvoxamine CR group versus placebo.[87] Also of note, there are 2 negative RCTs of fluoxetine in SAD.[88,89] However, Davidson *et al.*[90] compared fluoxetine (titrated to 40–60 mg/day), group comprehensive cognitive-behavioral therapy (CCBT), fluoxetine plus CCBT, and CCBT plus placebo versus placebo alone in a 14-week RCT and found significantly greater response versus placebo in all treatment groups by study end, with no differences between the active treatments. In another RCT comparing pharmacotherapy and psychotherapy in SAD, Blomhoff *et al.*[91] investigated sertraline monotherapy, exposure therapy, and combination treatment versus placebo or general medical care and found a greater percentage of responders in the sertraline group versus the non-sertraline group. Although only combination treatment resulted in significant response versus placebo beginning at week 12, both combination treatment and sertraline monotherapy resulted in significant response versus placebo by the end of the 24-week study.[91] There is also evidence from longer term relapse-prevention studies for paroxetine,[92] sertraline,[93] and escitalopram[94] in SAD.

Two 12-week, multicenter RCTs in SAD demonstrated that venlafaxine XR (75–225 mg/day) resulted in significant LSAS reductions beginning at week 4[95] or week 6.[96] Of note, both studies reported small but significant increases in the supine systolic (~0.6–2.3 mm Hg) and diastolic (~0.25–2.6 mm Hg) blood pressure, supine pulse rate (~1.3–4.9 beats per minute), and cholesterol levels (~0.3 mmol/L) in the venlafaxine XR groups.[95,96] There is also evidence for the efficacy of venlafaxine XR in a longer term (6-month), multicenter RCT.[97] Both fixed low-dose venlafaxine XR (75 mg/day) and flexible higher doses (150–225 mg/day) resulted in significant decreases in LSAS versus placebo beginning at week 4 and significant remission rates on LSAS by study end.[97]

In 2 head-to-head RCTs comparing venlafaxine XR (75–225 mg/day) and paroxetine (20–50 mg/day) in SAD, there were no significant differences in efficacy between the 2 active treatment groups after 12 weeks, as both resulted in significant decreases in LSAS and significant response rates versus placebo.[98,99] Another head-to-head, 24-week, placebo-controlled study found that escitalopram (5 and 20 mg/day) and paroxetine (20 mg/day) led to significant LSAS reductions versus placebo after 12 weeks, and in the observed case sample, all 3 escitalopram groups (5, 10, and 20 mg/day) and the paroxetine group (20 mg/day) experienced significant LSAS reductions by study end.[100] Although escitalopram (20 mg/day) resulted in significantly greater reductions than paroxetine (20 mg/day) in LSAS from week 16 onward,[100] these doses are not necessarily equivalent.

BZDs

There is evidence to support the use of BZDs in SAD in patients resistant to or unable to tolerate SSRIs (reviewed by Davidson[3] and Schneier[74]). However, Seedat and Stein[101] conducted a small, 10-week RCT of open-label paroxetine given in combination with clonazepam or placebo in SAD, and no significant differences between the 2 groups were found early or later in treatment. In addition, BZDs may be used "as needed" in performance-type or nongeneralized SAD, for which they are taken about 30 minutes prior to a performance (reviewed by Schneier[74]).

Anticonvulsants

One multicenter, 10-week RCT of pregabalin in SAD reported that pregabalin at 600 mg/day, but not at 150 mg/day, demonstrated efficacy in reducing LSAS scores.[102] Although pregabalin at 600 mg/day resulted in a significant percentage of responders on CGI-I, the LSAS response percentage was not significant.[102] Gabapentin, which is approved by the

FDA for partial seizures, neuropathic pain, and postherpetic neuralgia, has been used "off-label" in anxiety disorders, although RCT data have shown only a modest benefit (reviewed by Mathew *et al.*[103]). In an RCT for SAD, gabapentin led to significant reductions in LSAS after 14 weeks, but the LSAS response rate was only 32% and not significant versus placebo.[104]

Nongeneralized SAD or "Performance Anxiety"

Beta-blockers may be taken approximately 1 hour before a performance to help performance-type SAD (reviewed by Schneier[74]). For example, propranolol may be used at an initial dose of 10 mg, with a target dose of 10 to 40 mg, and used on an as-needed basis, although side effects include hypotension and bradycardia.[74]

SAD Meta-Analyses

There are a number of meta-analyses of RCTs supporting the use of SSRIs in SAD.[105,106] In a meta-analysis comparing SSRIs and non-SSRIs in SAD, no significant differences between medications or medication classes, including SSRIs, an MAOI, BZD, and gabapentin, were found, although the use of SSRIs as first-line treatments in SAD was supported on the basis of their tolerability, their ability to treat comorbid conditions, and the stability of the ES (LSAS ES for the SSRIs was 0.65).[107] The Cochrane Collaboration reviewed 37 RCTs of SSRIs, MAOIs, and other medications (including BZDs, buspirone, and gabapentin) in SAD and concluded that patients who received any medication were less likely than those given placebo to be nonresponders.[108] SSRIs were found to reduce symptom severity on LSAS, whereas MAOIs did not. The authors suggested that there may be a publication bias, suggesting that there might be more variation in medication responses in SAD than in the trials included in the analysis. However, maintenance trials demonstrated decreased relapse risk of medication versus placebo.[108]

Another meta-analysis compared psychological interventions, including cognitive restructuring and exposure, with pharmacological treatments in SAD, including SSRIs, BZDs, and MAOIs, in RCTs and non-RCTs.[109] On self-report measures, SSRIs and BZDs were found to be equally effective, with the largest mean ES compared to psychological and other pharmacological treatments (ES for SSRIs = 1.697, BZDs = 2.095), and their confidence intervals overlapped. A trend favored BZDs overall, which outperformed MAOIs and the psychological treatments, although SSRIs did not; on observer-rated measures, which included data from only a small number of trials, all of the treatments, including BZDs, SSRIs, MAOIs, and exposure with cognitive restructuring, were similarly effective and superior to placebo or wait-list control.[109]

POST-TRAUMATIC STRESS DISORDER

PTSD, which may occur after exposure to a traumatic event, is defined by the *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. (DSM-IV), by 3 symptom clusters–re-experiencing the event, emotional numbing/avoidance of stimuli associated with the trauma, and hyperarousal–that persist for more than 1 month and lead to functional impairment.[5] The "gold standard" in assessing PTSD symptoms is the Clinician-Administered PTSD Scale (CAPS), which is a structured interview that has been shown to be both reliable and valid.[110] CAPS is used in many RCTs to determine the efficacy of pharmacotherapy in PTSD.

As Mathew *et al.*[103] observed, a major limitation in the PTSD psychopharmacology literature is the lack of long-term studies (see the Cochrane Review of PTSD[4] discussed later). Because of limitations in the pharmacotherapy data for PTSD, guidelines from the National Institute for Health and Clinical Excellence (www.nice.org.uk), an independent source, have suggested that pharmacotherapy should not be used as a routine first-line PTSD treatment for adults, in preference to trauma-focused psychological therapy, on account of stronger evidence for exposure-based psychotherapies (reviewed by Mathew *et al.*[103]).

SSRIs and SNRIs

Two SSRIs (paroxetine and sertraline) have received FDA approval for PTSD, although in short-term trials ESs have been modest, and only paroxetine has been shown to be superior to placebo in all 3 PTSD symptom clusters.[111,112] A 6-month RCT of flexibly dosed venlafaxine XR (37.5–300 mg/day) in PTSD reported improvements in re-experiencing and avoidance/numbing but not in hyperarousal symptoms; remission rates (CAPS score ≤ 20) in the venlafaxine XR group were almost 51% and 37.5% of those in the placebo group.[113] Although there was overall efficacy versus placebo, the CAPS ES was small (about 0.3).[113] Davidson *et al.*[114] conducted a 12-week, multicenter RCT with placebo and sertraline control arms and reported efficacy of venlafaxine XR in PTSD, although the ES versus placebo was 0.266. Both of these studies used higher maximum daily venlafaxine XR doses than the FDA-approved doses for other indications, including major depression and other anxiety disorders.

Anticonvulsants

There have been very limited studies and few RCTs of anticonvulsants in PTSD, and none is FDA-approved for this indication. With respect to RCTs, there was a multicenter, adequately sized RCT of tiagabine in PTSD, which was negative, finding no efficacy of tiagabine versus placebo.[115] There was 1 very small, preliminary 12-week RCT of lamotrigine, which is FDA-approved for maintenance therapy in bipolar disorder, in 15 patients who met DSM-IV-TR criteria for a primary diagnosis of PTSD, which suggested some benefit.[116] There was 1 relatively small ($n = 38$) 12-week RCT of topiramate, which is approved by the FDA for migraine prophylaxis and seizures, in non–combat-related PTSD, which found no significant difference in CAPS scores between topiramate and placebo groups by study end.[117] There are only open-label and retrospective studies of other anticonvulsants in PTSD (reviewed by Mula *et al.*[37]).

Antipsychotics

There is some preliminary evidence for adjunctive risperidone in combat-related PTSD (reviewed by Gao *et al.*[118]). In 1 RCT in male patients with chronic, combat-related PTSD, adjunctive risperidone, added to a stable psychotropic regimen in most patients, led to significant CAPS improvement after 16 weeks, although treatment included an initial 5 weeks at a VA residential psychosocial program.[119] Another study in combat veterans with chronic PTSD and psychotic symptoms reported that adjunctive risperidone led to significant improvement in psychotic symptoms but not CAPS scores versus placebo after 5 weeks.[120] Padala *et al.*[121] conducted a small, 12-week pilot RCT of risperidone monotherapy in women who had PTSD due to sexual assault and domestic abuse, and they found that risperidone led to a decrease in PTSD symptoms. There are more inconsistent findings for olanzapine in PTSD, as there was 1 small negative RCT of olanzapine monotherapy,[122] although another small RCT of olanzapine augmentation of SSRIs found significant CAPS score reductions versus placebo after 8 weeks but no significant difference in the percentage of CGI-I responders, and patients in the olanzapine group gained approximately 13 lb on average.[123]

Alpha-1 Adrenergic Antagonists

As sleep disturbance occurs in up to 70% of patients with PTSD and norepinephrine is involved in stress-related arousal, the alpha-1-adrenergic antagonist prazosin, an FDA-approved antihypertensive and treatment for benign prostatic hypertrophy, has been investigated specifically to target nightmares in PTSD (reviewed by Dierks *et al.*[124]). There is preliminary evidence from relatively small RCTs and open-label trials and a retrospective chart review, which have their limitations, for prazosin treatment of PTSD-related nightmares, although further RCTs are indicated (reviewed by Dierks *et al.*[124]).

PTS Meta-Analysis

The Cochrane Collaboration also conducted a meta-analysis of 35 short-term (≤14 weeks) RCTs of pharmacotherapy of PTSD, finding that pharmacotherapy demonstrated efficacy in all PTSD symptom domains compared to placebo and in treating comorbid depressive symptoms and improving quality of life.[4] In the groups receiving medication, symptom severity and CGI-I response were significantly better in 17 and 13 trials, respectively.[4] Although there were limited direct-comparison trials of medications, SSRI trials contributed the most to overall ES and treatment response in the meta-analysis, and this led the authors to support the recommendation of SSRIs as first-line medications in PTSD.[4] There is also evidence for the efficacy of SSRIs in relapse prevention in 3 studies, which were reviewed by Stein *et al.*,[4] who reported that these results are consistent with recommendations for 6 to 12 months of medication treatment for acute PTSD and 12 months or more of interventions for chronic PTSD in order to prevent relapse. Notably, Stein *et al.* observed that war veterans are more resistant to medication treatment than other patients with PTSD, and there are few studies that directly compare psychotherapy and psychopharmacology in PTSD.

TCAs and MAOIs

TCAs are effective treatments for GAD and have a long history in treating anxiety, although few studies have investigated them in DSM-IV–defined GAD.[1] In an 8-week RCT of patients with DSM-III GAD by Rickels *et al.*[27] comparing imipramine, diazepam, and trazodone, diazepam resulted in the greatest improvement during the first 2 weeks, after which imipramine was significantly more effective in reducing total HAM-A scores (reviewed by Baldwin and Polkinghorn[125]); all 3 treatment groups were found to be effective in reducing anxiety symptoms by study end (reviewed by Lydiard and Monnier[126]). However, the use of TCAs is limited by their overall poorer tolerability in comparison with the newer SSRIs and SNRIs.

There is evidence from short- and long-term studies for the TCA clomipramine in PD.[50,51] One randomized, placebo-controlled trial compared patients with PD who received imipramine alone, cognitive-behavioral therapy (CBT) alone, CBT plus imipramine, or CBT plus placebo for an acute phase (3 months), and responders were followed for a maintenance phase (6 months) and another 6 months after they discontinued treatment.[127] In both the acute and maintenance phases, imipramine and CBT were each statistically superior to placebo on the PDSS.[127] In the acute phase, CBT plus imipramine was not superior to CBT plus placebo, but in the maintenance phase, the combined treatment led to significantly lower PDSS scores, but not PDSS or CGI-I response rates, in comparison with CBT plus placebo; after treatment discontinuation, CBT appeared to have a more lasting effect, with greater response rates in the CBT alone and CBT plus placebo groups versus placebo at 6-month follow-up.[127] Of note, there are no RCTs supporting TCAs in generalized SAD.

Although there is evidence for the efficacy of MAOIs in PD, SAD, and PTSD,[128] their use in these disorders is limited because of their poorer tolerability and risks of toxicity, as patients taking MAOIs are required to follow a low-tyramine diet to prevent hypertensive crisis.[3]

CONCLUSION

Evidence from multiple RCTs supports the use of SSRIs as first-line pharmacotherapy in the treatment of anxiety disorders, including GAD, PD, SAD, and PTSD. SNRIs, namely, venlafaxine XR and duloxetine, are now FDA-approved for the treatment of GAD, and venlafaxine XR is also approved by the FDA to treat PD and SAD, so SNRIs are emerging as first-line medications in anxiety disorders as well. Although older antidepressant agents, such as TCAs, also are efficacious in treating anxiety disorders, their use in clinical practice is limited by their less favorable side-effect profile in comparison with SSRIs; however, meta-analyses have demonstrated similar dropout rates for these 2 medication groups. The use of MAOIs is also limited by the need for patients to maintain a low-tyramine diet to prevent hypertensive crisis. BZDs are likewise very effective in treating anxiety disorders, and there are studies supporting their use as adjunctive agents in the short-term management of these disorders, given their rapid onset of action and ability to be used on an as-needed basis, whereas their long-term use is limited by the potential for abuse and dependence as well as their lack of antidepressant effects, as depression is often comorbid with the anxiety disorders. Pregabalin, an adjunctive anticonvulsant that is FDA-approved for treating postherpetic neuralgia, diabetic peripheral neuropathy, and fibromyalgia, is being increasingly studied in GAD, as there is preliminary evidence of efficacy, although long-term studies are needed. Buspirone remains an alternative treatment with no abuse liability in GAD, although it has not been found to be effective in PD or SAD. More recent studies are investigating the potential use of atypical antipsychotics in anxiety disorders because of their broad effects on the serotonin system, although further studies are needed. In PTSD, meta-analytic evidence supports the use of SSRIs as first-line pharmacotherapy, although National Institute for Health and Clinical Excellence guidelines have suggested that there is greater evidence to support trauma-focused psychotherapy over pharmacotherapy in PTSD. Newer agents, such as the alpha-1-antagonists, are being studied for their potential to treat nightmares in PTSD because of their noradrenergic effects. Meta-analyses continue to support the efficacy of both psychotherapy and pharmacotherapy in treating anxiety disorders. Although further studies of effectiveness data for the various anxiolytics are indicated, a recent STAR*D study found that patients with "anxious depression" were less likely to remit and more likely to develop side effects in response to pharmacotherapy, suggesting that comorbid anxiety contributes to poorer overall outcomes in real-world settings.[131]

DISCLOSURES

Dr. Mathew has received lecture or consulting fees from AstraZeneca, Jazz Pharmaceuticals, Pfizer, and Hoffman-LaRoche, and has received research support from Alexza Pharmaceuticals, Epix Pharmaceuticals, NARSAD, NIMH, and Novartis.

Dr. Mathew has been named as an inventor on a use-patent of ketamine for the treatment of depression. If ketamine were shown to be effective in the treatment of depression and received approval from the Food and Drug Administration (FDA) for this indication, Dr. Mathew could benefit financially.

Dr. Mathew is supported by NIMH grant K23-MH-069656.

REFERENCES

1. Mathew SJ, Hoffman EJ. Pharmacotherapy of generalized anxiety disorder. In: Antony M, Stein M,

eds. *Handbook of Anxiety and Anxiety Disorders*. New York, NY: Oxford University Press. In press.
2. Katon WJ. Panic disorder. *N Engl J Med* 2006; 354: 2360–2367.
3. Davidson JRT. Pharmacotherapy of social anxiety disorder: what does the evidence tell us? *J Clin Psychiatry* 2006; 67(suppl 12): 20–26.
4. Stein DJ, Ipser JC, Seedat S. Pharmacotherapy for post traumatic stress disorder (PTSD). *Cochrane Database Syst Rev* 2006; CD002795.
5. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Text Revision. Washington, DC: American Psychiatric Association; 2000.
6. Hamilton M. The assessment of anxiety states by rating. *Br J Med Psychol* 1959; 32: 50–55.
7. Guy W. Clinical global impression. In: *ECDEU Assessment Manual for Psychopharmacology*. Rev ed. Rockville, MD: National Institute of Mental Health; 1976.
8. Rickels K, Zaninelli R, McCafferty J, et al. Paroxetine treatment of generalized anxiety disorder: a double-blind, placebo-controlled study. *Am J Psychiatry* 2003; 160: 749–756.
9. Pollack MH, Zaninelli R, Goddard A, et al. Paroxetine in the treatment of generalized anxiety disorder: results of a placebo-controlled, flexible-dosage trial. *J Clin Psychiatry* 2001; 62: 350–357.
10. Davidson JRT, Bose A, Korotzer A, Zheng H. Escitalopram in the treatment of generalized anxiety disorder: double-blind, placebo controlled, flexible-dose study. *Depress Anxiety* 2004; 19: 234–240.
11. Davidson JRT, Bose A, Wang Q. Safety and efficacy of escitalopram in the long term treatment of generalized anxiety disorder. *J Clin Psychiatry* 2005; 66: 1441–1446.
12. Allgulander C, Florea I, Trap Huusom AK. Prevention of relapse in generalized anxiety disorder by escitalopram treatment. *Int J Neuropsychopharmacol* 2005; 9: 1–11.
13. Stocchi F, Nordera G, Jokinen RH, et al., for the Paroxetine Generalized Anxiety Disorder Study Team. Efficacy and tolerability of paroxetine for the long-term treatment of generalized anxiety disorder. *J Clin Psychiatry* 2003; 64: 250–258.
14. Dahl AA, Ravindran A, Allgulander C, et al. Sertraline in generalized anxiety disorder: efficacy in treating the psychic and somatic anxiety factors. *Acta Psychiatr Scand* 2005; 111: 429–435.
15. Allgulander C, Dahl AA, Austin C, et al. Efficacy of sertraline in a 12-week trial for generalized anxiety disorder. *Am J Psychiatry* 2004; 161: 1642–1649.
16. Allgulander C, Hackett D, Salinas E. Venlafaxine extended release (ER) in the treatment of generalised anxiety disorder: twenty-four-week placebo-controlled dose-ranging study. *Br J Psychiatry* 2001; 179: 15–22.
17. Gelenberg AJ, Lydiard RB, Rudolph RL, et al. Efficacy of venlafaxine extended-release capsules in nondepressed outpatients with generalized anxiety disorder: a 6-month randomized controlled trial. *JAMA* 2000; 283: 3082–3088.
18. Rickels K, Pollack MH, Sheehan DV, Haskins JT. Efficacy of extended-release venlafaxine in nondepressed outpatients with generalized anxiety disorder. *Am J Psychiatry* 2000; 157: 968–974.
19. Davidson JRT, DuPont RL, Hedges D, Haskins JT. Efficacy, safety, and tolerability of venlafaxine extended release and buspirone in outpatients with generalized anxiety disorder. *J Clin Psychiatry* 1999; 60: 528–535.
20. Nimatoudis I, Zissis NP, Kogeorgos J, et al. Remission rates with venlafaxine extended release in Greek outpatients with generalized anxiety disorder. A double-blind, randomized, placebo controlled study. *Int Clin Psychopharmacol* 2004; 19: 331–336.
21. Hackett D, Haudiquet V, Salinas E. A method for controlling for a high placebo response rate in a comparison of venlafaxine XR and diazepam in the short-term treatment of patients with generalised anxiety disorder. *Eur Psychiatry* 2003; 18: 182–187.
22. Rynn M, Russell J, Erickson J, et al. Efficacy and safety of duloxetine in the treatment of generalized anxiety disorder: a flexible-dose, progressive-titration, placebo-controlled trial. *Depress Anxiety* 2008; 25: 182–189.
23. Hartford J, Kornstein S, Liebowitz M, et al. Duloxetine as an SNRI treatment for generalized anxiety disorder: results from a placebo and active-controlled trial. *Int Clin Psychopharmacol* 2007; 22: 167–174.
24. Bielski RJ, Bose A, Chang, C-C. A double-blind comparison of escitalopram and paroxetine in the long-term treatment of generalized anxiety disorder. *Ann Clin Psychiatry* 2005; 17: 65–69.
25. Ball SG, Kuhn A, Wall D, et al. Selective serotonin reuptake inhibitor treatment for generalized anxiety disorder: a double-blind, prospective comparison between paroxetine and sertraline. *J Clin Psychiatry* 2005; 66: 94–99.
26. Kim TS, Pae CU, Yoon SJ, et al. Comparison of venlafaxine extended release versus paroxetine for treatment of patients with generalized anxiety disorder. *Psychiatry Clin Neurosci* 2006; 60: 347–351.
27. Rickels K, Downing R, Schweizer E, Hassman H. Antidepressants for the treatment of generalized anxiety disorder: a placebo-controlled comparison of imipramine, trazodone, and diazepam. *Arch Gen Psychiatry* 1993; 50: 884–895.
28. Rocca P, Fonzo V, Scotta M, et al. Paroxetine efficacy in the treatment of generalized anxiety disorder. *Acta Psychiatr Scand* 1997; 95: 444–450.
29. Gorman JM. Treating generalized anxiety disorder. *J Clin Psychiatry* 2003; 64(suppl 2): 24–29.
30. O'Brien CP. Benzodiazepine use, abuse, and dependence. *J Clin Psychiatry* 2005; 66(suppl 2): 28–33.
31. Gammans RE, Stringfellow JC, Hvizdos AJ, et al. Use of buspirone in patients with generalized anxiety disorder and coexisting depressive symptoms. A meta-analysis of eight, randomized, controlled studies. *Neuropsychobiology* 1992; 25: 193–201.
32. DeMartinis N, Rynn M, Rickels K, Mandos L. Prior benzodiazepine use and buspirone response in the treatment of generalized anxiety disorder. *J Clin Psychiatry* 2000; 62: 657–658.
33. Sheehan DV, Raj AB, Harnett-Sheehan K, et al. The relative efficacy of high dose buspirone and alprazolam in the treatment of panic disorder: a double-blind placebo-controlled study. *Acta Psychiatr Scand* 1993; 88: 1–11.

34. Sheehan DV, Raj AB, Harnett-Sheehan K, Soto S. Is buspirone effective for panic disorder? *J Clin Psychopharmacol* 1993; 10: 3–11.
35. Pohl R, Balon R, Yeragani VK, Gershon S. Serotonergic anxiolytics in the treatment of panic disorder: a controlled study with buspirone. *Psychopathology* 1989; 22(suppl 1): 60–67.
36. van Vliet IM, den Boer JA, Westenberg HG, Pian KL. Clinical effects of buspirone in social phobia: a double-blind placebo-controlled study. *J Clin Psychiatry* 1997; 58: 164–168.
37. Mula M, Pini S, Cassano GB. The role of anticonvulsant drugs in anxiety disorders: a critical review of the evidence. *J Clin Psychopharmacol* 2007; 27: 263–272.
38. Bech P. Dose-response relationship of pregabalin in patients with generalized anxiety disorder. A pooled analysis of four placebo-controlled trials. *Pharmacopsychiatry* 2007; 40: 163–168.
39. Feltner DE, Crockatt JG, Dubovsky SJ, et al. A randomized, double-blind, placebo-controlled, fixed-dose, multicenter study of pregabalin in patients with generalized anxiety disorder. *J Clin Psychopharmacol* 2003; 223: 240–249.
40. Pande AC, Crockatt JG, Feltner DE, et al. Pregabalin in generalized anxiety disorder: a placebo-controlled trial. *Am J Psychiatry* 2003; 160: 533–540.
41. Pohl RB, Feltner DE, Fieve RR, Pande AC. Efficacy of pregabalin in the treatment of generalized anxiety disorder. *J Clin Psychopharmacol* 2005; 25: 151–158.
42. Rickels K, Pollack MH, Feltner DE, et al. Pregabalin for treatment of generalized anxiety disorder: a 4-week, multicenter, double-blind, placebo-controlled trial of pregabalin and alprazolam. *Arch Gen Psychiatry* 2005; 62: 1022–1030.
43. Montgomery SA, Tobias K, Zornberg GL, et al. Efficacy and safety of pregabalin in the treatment of generalized anxiety disorder: a 6-week, multicenter, randomized, double-blind placebo-controlled comparison of pregabalin and venlafaxine. *J Clin Psychiatry* 2006; 67: 771–782.
44. Pollack MJ, Simon NM, Zalta AK, et al. Olanzapine augmentation of fluoxetine for refractory generalized anxiety disorder: a placebo controlled study. *Biol Psychiatry* 2006; 59: 211–215.
45. Brawman-Mintzer O, Knapp RG, Nietert PJ. Adjunctive risperidone in generalized anxiety disorder: a double-blind, placebo-controlled study. *J Clin Psychiatry* 2005; 66: 1321–1325.
46. Kapczinski F, Lima MS, Souza JS, et al. Antidepressants for generalized anxiety disorder. *Cochrane Database Syst Rev* 2003; CD003592.
47. Shear MK, Brown TA, Barlow DH, et al. Multicenter collaborative panic disorder severity scale. *Am J Psychiatry* 1997; 154: 11.
48. Huppert JD, Schultz LT, Foa EB, et al. Differential response to placebo among patients with social phobia, panic disorder, and obsessive-compulsive disorder. *Am J Psychiatry* 2004; 161: 1485–1487.
49. Heuer L, Mathew SJ, Charney DS. Panic disorder. In: Squire L, Albright T, Bloom F, et al., eds. *New Encyclopedia of Neuroscience*. Oxford, England: Elsevier, Ltd. In press.
50. Lecrubier Y, Bakker A, Dunbar G, Judge R. A comparison of paroxetine, clomipramine and placebo in the treatment of panic disorder. Collaborative Paroxetine Panic Study Investigators. *Acta Psychiatr Scand* 1997; 95: 145–152.
51. Lecrubier Y, Judge R. Long-term evaluation of paroxetine, clomipramine and placebo in panic disorder. Collaborative Paroxetine Panic Study Investigators. *Acta Psychiatr Scand* 1997; 95: 153–160.
52. Ballenger JC, Wheadon DE, Steiner M, et al. Double-blind, fixed-dose, placebo-controlled study of paroxetine in the treatment of panic disorder. *Am J Psychiatry* 1998; 155: 36–42.
53. Sheehan DV, Burnham DB, Iyengar MK, et al. Efficacy and tolerability of controlled-release paroxetine in the treatment of panic disorder. *J Clin Psychiatry* 2005; 66: 34–40.
54. Pollack MH, Otto MW, Worthington JJ, et al. Sertraline in the treatment of panic disorder. *Arch Gen Psychiatry* 1998; 55: 1010–1016.
55. Pohl RB, Wolkow RM, Clary CM. Sertraline in the treatment of panic disorder: a double-blind multicenter trial. *Am J Psychiatry* 1998; 155: 1189–1195.
56. Rapaport MH, Wolkow R, Rubin A, et al. Sertraline treatment of panic disorder: results of a long-term study. *Acta Psychiatr Scand* 2001; 104: 289–298.
57. Michelson D, Allgulander K, Dantendorfer K, et al. Efficacy of usual antidepressant dosing regimens of fluoxetine in panic disorder. *Br J Psychiatry* 2001; 179: 514–518.
58. Asnis GM, Hameedi FA, Goddard AW, et al. Fluvoxamine in the treatment of panic disorder: a multi-center, double-blind, placebo-controlled study in outpatients. *Psychiatry Res* 2001; 103: 1–14.
59. Hoehn-Saric R, McLeod DR, Hipsley PA. Effect of fluvoxamine on panic disorder. *J Clin Psychopharmacol* 1993; 13: 321–326.
60. Black DW, Wesner R, Bowers W, Gabel J. A comparison of fluvoxamine, cognitive therapy, and placebo in the treatment of panic disorder. *Arch Gen Psychiatry* 1993; 50: 44–50.
61. Stahl SM, Gergel I, Li D. Escitalopram in the treatment of panic disorder: a randomized, double-blind, placebo-controlled trial. *J Clin Psychiatry* 2003; 64: 1322–1327.
62. Wade AG, Lepola U, Koponen HJ, et al. The effect of citalopram in panic disorder. *Br J Psychiatry* 1997; 170: 549–553.
63. Leopola UM, Wade AG, Leinonen EV, et al. A controlled, prospective, 1-year trial of citalopram in the treatment of panic disorder. *J Clin Psychiatry* 1998; 59: 528–534.
64. Michelson D, Lydiard RB, Pollack MH, et al. Outcome assessment and clinical improvement in panic disorder: evidence from a randomized controlled trial of fluoxetine and placebo. *Am J Psychiatry* 1998; 155: 1570–1577.
65. Bandelow B, Behnke K, Lenoir S, et al. Sertraline versus paroxetine in the treatment of panic disorder: an acute, double-blind noninferiority comparison. *J Clin Psychiatry* 2004; 65: 405–413.
66. Pollack MH, Lepola U, Koponen H, et al. A double-blind study of the efficacy of venlafaxine extended-release, paroxetine, and placebo in the treatment of panic disorder. *Depress Anxiety* 2007; 24: 1–14.

67. Bradwejn J, Ahokas A, Stein DJ, et al. Venlafaxine extended-release capsules in panic disorder: flexible-dose, double-blind, placebo-controlled study. *Br J Psychiatry* 2005; 187: 352–359.
68. Ferguson JM, Khan A, Mangano R, et al. Relapse prevention of panic disorder in adult outpatient responders to treatment with venlafaxine extended release. *J Clin Psychiatry* 2007; 68: 58–68.
69. Goddard AW, Brouette T, Almai A, et al. Early coadministration of clonazepam with sertraline for panic disorder. *Arch Gen Psychiatry* 2001; 58: 681–686.
70. Pollack MH, Simon NM, Worthington JJ, et al. Combined paroxetine and clonazepam treatment strategies compared to paroxetine monotherapy for panic disorder. *J Psychopharmacol* 2003; 17: 276–282.
71. Mitte K. A meta-analysis of the efficacy of psycho- and pharmacotherapy in panic disorder with and without agoraphobia. *J Affect Disord* 2005; 88: 27–45.
72. Otto MW, Tuby KS, Gould RA, et al. An effect-size analysis of the relative efficacy and tolerability of serotonin selective reuptake inhibitors for panic disorder. *Am J Psychiatry* 2001; 158: 1989–1992.
73. Furukawa TA, Watanabe N, Churchill R. Combined psychotherapy plus antidepressants for panic disorder with or without agoraphobia. *Cochrane Database Syst Rev* 2007; CD004364.
74. Schneier FR. Social anxiety disorder. *N Engl J Med* 2006; 355: 1029–1036.
75. Liebowitz MR. Social phobia. *Mod Probl Pharmacopsychiatry* 1987; 22: 141–173.
76. Stein MB, Liebowitz MR, Lydiard B, et al. Paroxetine treatment of generalized social phobia (social anxiety disorder): a randomized controlled trial. *JAMA* 1998; 280: 708–713.
77. Baldwin D, Bobes J, Stein DJ, et al., for the Paroxetine Study Group. Paroxetine in social phobia/social anxiety disorder. Randomised, double-blind, placebo-controlled study. *Br J Psychiatry* 1999; 175: 120–126.
78. Lepola U, Bergtholdt B, St. Lambert J, et al. Controlled-release paroxetine in the treatment of patients with social anxiety disorder. *J Clin Psychiatry* 2004; 65: 222–229.
79. Liebowitz MR, Stein MB, Tancer M, et al. A randomized, double-blind, fixed-dose comparison of paroxetine and placebo in the treatment of generalized social anxiety disorder. *J Clin Psychiatry* 2002; 63: 66–74.
80. Liebowitz MR, DeMartinis NA, Weihs K, et al. Efficacy of sertraline in severe generalized social anxiety disorder: results of a double-blind, placebo-controlled study. *J Clin Psychiatry* 2003; 64: 785–792.
81. Van Ameringen MA, Lane RM, Walker J, et al. Sertraline treatment of generalized social phobia: a 20-week, double-blind, placebo-controlled study. *Am J Psychiatry* 2001; 158: 275–281.
82. Kasper S, Stein DJ, Loft H, Nil R. Escitalopram in the treatment of social anxiety disorder. *Br J Psychiatry* 2005; 186: 222–226.
83. Stein MB, Fyer AJ, Davidson JRT, et al. Fluvoxamine treatment of social phobia (social anxiety disorder): a double-blind, placebo-controlled study. *Am J Psychiatry* 1999; 156: 756–760.
84. Asakura S, Tajima O, Koyama T. Fluvoxamine treatment of generalized social anxiety disorder in Japan: a randomized double-blind, placebo-controlled study. *Int J Neuropsychopharmacol* 2007; 10: 263–274.
85. Westenberg GM, Stein DJ, Yang H, et al. A double-blind placebo-controlled study of controlled release fluvoxamine for the treatment of generalized social anxiety disorder. *J Clin Psychopharmacol* 2004; 24: 49–55.
86. Davidson J, Yaryura-Tobias J, DuPont R, et al. Fluvoxamine-controlled release formulation for the treatment of generalized social anxiety disorder. *J Clin Psychopharmacol* 2004; 24: 118–125.
87. Stein DJ, Westenberg HGM, Yang H, et al. Fluvoxamine CR in the long-term treatment of social anxiety disorder: the 12- to 24-week extension phase of a multicentre, randomized, placebo-controlled trial. *Int J Neuropsychopharmacol* 2003; 6: 317–323.
88. Clark DM, Ehlers A, McManus F, et al. Cognitive therapy versus floxetine in generalized social phobia: a randomized placebo-controlled trial. *J Consult Clin Psychol* 2003; 71: 1058–1067.
89. Kobak KA, Greist JH, Jefferson JW, Katzelnick DJ. Fluoxetine in social phobia: a double-blind, placebo-controlled pilot study. *J Clin Psychopharmacol* 2002; 22: 257–262.
90. Davidson JRT, Foa EB, Huppert JD, et al. Fluoxetine, comprehensive cognitive behavioral therapy, and placebo in generalized social phobia. *Arch Gen Psychiatry* 2004; 61: 1005–1013.
91. Blomhoff S, Haug TT, Hellström K, et al. Randomised controlled general practice trial of sertraline, exposure therapy and combined treatment in generalised social phobia. *Br J Psychiatry* 2001; 179: 23–30.
92. Stein DJ, Versiani M, Hair T, Kumar R. Efficacy of paroxetine for relapse prevention in social anxiety disorder: a 24-week study. *Arch Gen Psychiatry* 2002; 59: 1111–1118.
93. Walker JR, Van Ameringen MA, Swinson R, et al. Prevention of relapse in generalized social phobia: results of a 24-week study in responders to 20 weeks of sertraline treatment. *J Clin Psychopharmacol* 2000; 20: 636–644.
94. Montgomery SA, Nil R, Dürr-Pal N, et al. A 24-week randomized, double-blind, placebo-controlled study of escitalopram for the prevention of generalized social anxiety disorder. *J Clin Psychiatry* 2005; 66: 1270–1278.
95. Rickels K, Mangano, Khan A. A double-blind, placebo-controlled study of a flexible dose of venlafaxine ER in adult outpatients with generalized social anxiety disorder. *J Clin Psychopharmacol* 2004; 24: 488–496.
96. Liebowitz MR, Mangano RM, Bradwejn J, Asnis G. A randomized controlled trial of venlafaxine extended release in generalized social anxiety disorder. *J Clin Psychiatry* 2005; 66: 238–247.
97. Stein MB, Pollack MH, Bystritsky A, et al. Efficacy of low and higher dose extended-release venlafaxine in generalized social anxiety disorder: a 6-month randomized controlled trial. *Psychopharmacology* 2005; 177: 280–288.
98. Allgulander C, Mangano R, Zhang J, et al. Efficacy of venlafaxine ER in patients with social anxiety disorder: a double-blind, placebo-controlled, parallel-group comparison with paroxetine. *Hum Psychopharmacol* 2004; 19: 387–396.

99. Liebowitz MR, Gelenberg AJ, Munjack D. Venlafaxine extended release vs. placebo and paroxetine in social anxiety disorder. *Arch Gen Psychiatry* 2005; 62: 190–198.
100. Lader M, Stender K, Bürger V, Nil R. The efficacy and tolerability of escitalopram in the short- and long-term treatment of social anxiety disorder: a randomized, double-blind, placebo-controlled, fixed-dose study. *Depress Anxiety* 2004; 19: 241–248.
101. Seedat S, Stein MB. Double-blind placebo-controlled assessment of combined clonazepam with paroxetine compared with paroxetine monotherapy for generalized social anxiety disorder. *J Clin Psychiatry* 2004; 65: 244–248.
102. Pande AC, Feltner DE, Jefferson JW. Efficacy of the novel anxiolytic pregabalin in social anxiety disorder: a placebo-controlled, multicenter study. *J Clin Psychopharmacol* 2004; 24: 141–149.
103. Mathew SJ, Hoffman EJ, Charney DS. Pharmacotherapy of anxiety disorders. In: Charney D, Nestler EJ, eds. *Neurobiology of Mental Illness*. 3rd ed. New York, NY: Oxford University Press. In press.
104. Pande AC, Davidson JRT, Jefferson JW, et al. Treatment of social phobia with gabapentin: a placebo-controlled study. *J Clin Psychopharmacol* 1999; 19: 341–348.
105. Hedges DW, Brown BL, Shwalb DA, et al. The efficacy of selective serotonin reuptake inhibitors in adult social anxiety disorder: a meta-analysis of double-blind, placebo-controlled trials. *J Psychopharmacol* 2007; 21: 102–111.
106. Van der Linden GJ, Stein DJ, van Balkom AJ. The efficacy of the selective serotonin reuptake inhibitors for social anxiety disorder (social phobia): a meta-analysis of randomized controlled trials. *Int Clin Psychopharmacol* 2000; 15(suppl 2): S15–S23.
107. Blanco C, Schneier FR, Schmidt A, et al. Pharmacological treatment of social anxiety disorder: a meta-analysis. *Depress Anxiety* 2003; 18: 29–40.
108. Stein DJ, Ipser JC, van Balkom AJ. Pharmacotherapy for social anxiety disorder. *Cochrane Database Syst Rev* 2000; CD001206.
109. Federoff IC, Taylor S. Psychological and pharmacological treatments of social phobia: a meta-analysis. *J Clin Psychopharmacol* 2001; 21: 311–324.
110. Blake DD, Weathers FW, Nagy LM, et al. The development of a clinician-administered PTSD scale. *J Trauma Stress* 1995; 8: 75–90.
111. Tucker P, Zaninelli R, Yehuda R, et al. Paroxetine in the treatment of chronic posttraumatic disorder: results of a placebo-controlled, flexible-dosage trial. *J Clin Psychiatry* 2001; 62: 860–868.
112. Marshall RD, Beebe KL, Oldham M, Zaninelli R. Efficacy and safety of paroxetine treatment for chronic PTSD: a fixed-dose, placebo-controlled study. *Am J Psychiatry* 2001; 158: 1982–1988.
113. Davidson J, Baldwin D, Stein DJ, et al. Treatment of posttraumatic stress disorder with venlafaxine extended release: a 6-month randomized controlled trial. *Arch Gen Psychiatry* 2006; 63: 1158–1165.
114. Davidson J, Rothbaum BO, Tucker P, et al. Venlafaxine extended release in posttraumatic stress disorder: a sertraline- and placebo-controlled study. *J Clin Psychopharmacol* 2006; 26: 259–267.
115. Davidson JRT, Brady K, Mellman TA, et al. The efficacy and tolerability of tiagabine in adult patients with post-traumatic stress disorder. *J Clin Psychopharmacol* 2007; 27: 85–88.
116. Hertzberg MA, Butterfield MI, Feldman ME, et al. A preliminary study of lamotrigine for the treatment of posttraumatic stress disorder. *Biol Psychiatry* 1999; 45: 1226–1229.
117. Tucker P, Trautman RP, Wyatt DB, et al. Efficacy and safety of topiramate monotherapy in civilian posttraumatic stress disorder: a randomized, double-blind, placebo-controlled study. *J Clin Psychiatry* 2007; 68: 201–206.
118. Gao K, Muzina D, Gajwani P, Calabrese JR. Efficacy of typical and atypical antipsychotics for primary and comorbid anxiety symptoms or disorders: a review. *J Clin Psychiatry* 2006; 67: 1327–1340.
119. Bartzokis G, Lu PH, Turner J, et al. Adjunctive risperidone in the treatment of chronic combat-related posttraumatic stress disorder. *Biol Psychiatry* 2004; 57: 474–479.
120. Hamner MB, Faldowski RA, Ulmer HG, et al. Adjunctive risperidone treatment in post-traumatic stress disorder: a preliminary controlled trial of effects on comorbid psychotic symptoms. *Int Clin Psychopharmacol* 2003; 18: 1–8.
121. Padala PR, Madison J, Monnahan M, et al. Risperidone monotherapy for post-traumatic stress disorder related to sexual assault and domestic abuse in women. *Int Clin Psychopharmacol* 2006; 21: 275–280.
122. Butterfield MI, Becker ME, Connor KM, et al. Olanzapine in the treatment of post-traumatic stress disorder: a pilot study. *Int Clin Psychopharmacol* 2001; 16: 197–203.
123. Stein MB, Kline NA, Matloff JL. Adjunctive olanzapine for SSRI-resistant combat-related PTSD: a double-blind, placebo-controlled study. *Am J Psychiatry* 2002; 159: 1777–1779.
124. Dierks MR, Jordan JK, Sheehan AH. Prazosin treatment of nightmares related to posttraumatic stress disorder. *Ann Pharmacother* 2007; 41: 1013–1017.
125. Baldwin DS, Polkinghorn C. Evidence-based pharmacotherapy of generalized anxiety disorder. *Int J Neuropsychopharmacol* 2005; 8: 293–302.
126. Lydiard RB, Monnier J. Pharmacological treatment. In: Heimberg RC, Turk CL, Mannin DS, editors. *Generalized Anxiety Disorder: Advances in Research and Practice*. New York, NY: Guilford Press; 2004; 351–379.
127. Barlow DH, Gorman JM, Shear MK, Woods SW. Cognitive-behavioral therapy, imipramine, or their combination for panic disorder: a randomized controlled trial. *JAMA* 2000; 283: 2529–2536.
128. Liebowitz MR, Hollander E, Schneier F, et al. Reversible and irreversible monoamine oxidase inhibitors in other psychiatric disorders. *Acta Psychiatr Scand Suppl* 1990; 360: 29–34.
129. Rosenbaum JF, Arana GW, Hyman SE, et al. *Handbook of Psychiatric Drug Therapy*. 5th ed. Philadelphia, PA: Lippincott Williams&Wilkins; 2005.
130. Goddard AW, Coplan JD, Shekhar A, et al. Principles of pharmacotherapy for the anxiety disorders. In: Charney DS, ed. *Neurobiology of Mental Illness*. 2nd ed. New York, NY: Oxford University Press; 2002.
131. Fava M, Rush AJ, Alpert JE, et al. Difference in treatment outcome in outpatients with anxious versus nonanxious depression: a STAR*D report. *Am J Psychiatry* 2008 165: 342–351.

Neurostimulatory Therapeutics in Management of Treatment-Resistant Depression with Focus on Deep Brain Stimulation

Dani Dumitriu,[1] Kate Collins, MSW,[2] Ron Alterman, MD,[3]
and Sanjay J. Mathew, MD[2]

[1] Department of Neuroscience, Mount Sinai School of Medicine, New York, NY
[2] Department of Psychiatry, Mount Sinai School of Medicine, New York, NY
[3] Department of Neurosurgery, Mount Sinai School of Medicine, New York, NY

ABSTRACT

Treatment-resistant depression continues to pose a major medical challenge, as up to one-third of patients with major depressive disorder fail to have an adequate response to standard pharmacotherapies. An improved understanding of the complex circuitry underlying depressive disorders has fostered an explosion in the development of new, nonpharmacological approaches. Each of these treatments seeks to restore normal brain activity via electrical or magnetic stimulation. In this article, the authors discuss the ongoing evolution of neurostimulatory treatments for treatment-resistant depression, reviewing the methods, efficacy, and current research on electroconvulsive therapy, repetitive transcranial magnetic stimulation, magnetic seizure therapy, focal electrically administered stimulated seizure therapy, transcranial direct current stimulation, chronic epidural cortical stimulation, and vagus nerve stimulation. Special attention is given to deep brain stimulation, the most focally targeted approach. The history, purported mechanisms of action, and current research are outlined in detail. Although deep brain stimulation is the most invasive of the neurostimulatory treatments developed to date, it may hold significant promise in alleviating symptoms and improving the quality of life for patients with the most severe and disabling mood disorders. ***Mt Sinai J Med 75:263–275, 2008.*** © *2008 Mount Sinai School of Medicine*

Address Correspondence to:

Dani Dumitriu
Department of Neuroscience
Mount Sinai School of Medicine
New York, NY
Email: dani.dumitriu@mssm.edu

Key Words: cingulate gyrus, deep brain stimulation, electroconvulsive therapy, internal capsule, major depression, neurosurgery, nucleus accumbens, therapeutics, transcranial magnetic stimulation, treatment-resistant depression, vagal nerve stimulation, ventral striatum.

In 1957, a medication developed to treat tuberculosis fundamentally changed psychiatry when it was observed to have as one of its side effects mood improvement. Iproniazide, the first of the monoamine oxidase inhibitors, ushered in a new era in which innovations in biological treatments for depression were almost exclusively pharmacological. Furthermore, post hoc analyses of its mechanism spawned the monoamine theory of depression, which posits that low levels of monoamines precipitate depression.[1] This paradigm would dominate research aimed at developing novel treatments for depression by investigators in both academia and industry for almost half a century. The monoamine oxidase inhibitors, followed shortly by the tricyclic antidepressants, were generally effective in the treatment of depressive disorders; however, their side effects were prohibitive for many patients. The introduction of selective serotonin reuptake inhibitors (SSRIs) in the late 1980s, which are generally safer and easier to tolerate, precipitated optimism among clinicians treating mood disorders. The psychopharmacological boom that followed in the 1990s, with blockbuster drugs such as fluoxetine (Prozac), sertraline (Zoloft), and paroxetine (Paxil), led some to

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20044

© 2008 Mount Sinai School of Medicine

believe that nonpharmacological somatic interventions, such as electroconvulsive therapy (ECT), were obsolete.

Twenty years later, SSRIs are still considered the first-line treatment for depression. However, it has become clear that they and other medications with monoaminergic targets (eg, SNRIs, atypical antipsychotics, and bupropion) are hardly panaceas. In the largest clinical trial conducted in real-world patients with depression, the Sequenced Treatment Alternatives To Relieve Depression (STAR*D) trial, patients diagnosed with nonpsychotic major depressive disorder received 1 (n = 3671) to 4 (n = 123) consecutive treatments with antidepressant medications: 64.2% of patients failed to achieve remission after a vigorously dosed 10- to 12-week trial of citalopram.[2] Furthermore, there is evidence suggesting that the gains achieved from SSRIs and other antidepressants are lost once the medication is discontinued, and relapse and recurrence of illness are the norm. Specific forms of manualized psychotherapy such as cognitive behavioral therapy may in fact be more durable than SSRI treatment in longer term follow-up.[3]

In the past 2 decades, a new body of scholarship has emerged to address the failure of monoaminergic medications to adequately treat up to one-third of patients. Two primary treatment-resistant depression (TRD) research paradigms have emerged. The first seeks drugs with novel mechanisms and is discussed elsewhere in this issue (see Sanacora *et al.*). In this review, we address the second approach, which focuses on neurostimulatory interventions. The present status of ECT, repetitive transcranial magnetic stimulation (rTMS), magnetic seizure therapy (MST), focal electrically administered stimulated seizure therapy, transcranial direct current stimulation, chronic epidural cortical stimulation, vagus nerve stimulation (VNS), and deep brain stimulation (DBS) research is reviewed, and the implications for clinical practice and future investigation are discussed. Particular attention is given to DBS, potentially the most promising of these treatment modalities.

OVERVIEW OF TREATMENT-RESISTANT DEPRESSION

To date, the research on TRD has been plagued by methodological inconsistencies across research groups. Even its exact definition is still controversial: although TRD can be generally understood as a depression that fails to respond to at least 2 adequate trials of antidepressant therapy, there is no consensus regarding the criteria for adequacy. Only 1 empirically validated instrument measures the clinical strength of a patient's past medication trials. The Antidepressant Treatment History Questionnaire assesses a score from 0 to 5, reflecting the robustness of a given medication trial on the basis of the dosage and duration of treatment.[4–6] However, the criteria of the Antidepressant Treatment History Questionnaire are based on broad standards of practice and fail to account for individual variations in drug metabolism or treatment compliance (see Rush *et al.*[7]).

Furthermore, multiple staging paradigms persist, each defining the severity of treatment resistance with a different algorithm. Three of the most prominent models are the model developed by Thase and Rush,[8] the Massachusetts General Hospital Treatment-Resistant Scale,[9] and the European model.[10] The models' designation for duration of treatment, number of failed drug and ECT trials, and number of stages is highly variable. For example, stage 0 resistance in the European model corresponds closest to Thase and Rush's stage 1, and there is no corresponding stage in the European model for what Thase and Rush have assigned to stage 0: "all previous medication trials are inadequate." Thase and Rush's system is likely the most well known, but it has been criticized for assuming a hierarchy among different classes of antidepressants even though the empirical evidence may not support this assumption. The Massachusetts General Hospital scale does not use a staging paradigm; it uses a continuous variable, with higher scores indicating higher resistance.

The clinical significance of an individual patient's level of resistance, defined by any of the staging paradigms, is still unclear. A recent investigation of the clinical correlates of TRD revealed that nonresponse to the first antidepressant treatment was associated with future treatment-resistant status.[11] However, of 11 related factors, original nonresponse was the least potent predictor. Comorbid anxiety, panic disorder, social phobia, and personality disorders, as well as suicidal risk, depression severity, number of hospitalizations, early age at onset, and melancholic subtype, were more significantly associated with TRD.

Whatever the etiology of TRD is, it is clear that its prevalence is crippling both to the patient and to society as a major economic burden. Depression is a major cause of worldwide disability, with a lifetime risk of 16.2%.[12] Of these patients, up to one-third fail to respond to adequate drug therapy.[2,9] These treatment-resistant patients, representing a significant portion of our population, are often unable to work, hold meaningful social relationships, or even care for themselves. It has become clear that such a debilitating disease needs significant attention from the medical and research community.

ELECTROCONVULSIVE THERAPY

Traditionally reserved for TRD patients in the United States, ECT is the oldest continuously practiced somatic intervention in psychiatry. Over time, the methods of ECT administration have been increasingly refined. The putative therapeutic component, the elicitation of a seizure via conduction of electricity through the brain, has remained constant. Pioneered in 1938, ECT originally involved the delivery of energy to the brain via 2 electrodes placed on both temples.[13] Today, the bilateral method is still employed, but a unilateral form has also been developed with the aim of minimizing cognitive side effects. In unilateral ECT, 1 electrode is located on the nondominant temple, whereas the other is placed near the top of the head (vertex) on the same side. Other modern improvements include the utilization of anesthesia and muscle relaxants as well as the development of more sophisticated devices. Today, ECT remains the most effective intervention available; double-blinded, randomized, controlled studies comparing ECT with sham consistently have reported a 50% or greater response rate, and studies with more rigorous treatment protocols have reported response rates approaching 90%.[14]

Despite the indisputable evidence documenting the efficacy of ECT, few patients actually undergo the procedure: only about 100,000 treatments are performed annually in the United States versus an estimated 118 million patients for whom physicians prescribe antidepressants.[15,16] Why is the utilization of ECT so minimal? The best explanation at present is probably the stigma associated with this treatment (see Dowman *et al.*[17] for a review). However, ECT has some well-documented limitations: as many as 77% of patients experience relapse 6 months after therapy.[14] Cognitive side effects are also a major concern, although the data are less conclusive. Short-term memory loss and confusion are often observed in the immediate aftermath of treatment.[18] However, a recent review of ECT practice and success at a set of community clinics found that cognitive outcomes varied across treatment facilities and that differences in ECT technique largely accounted for this variance.[19]

An important explanation for the underuse of ECT may be the lack of ECT training offered to medical professionals. Psychiatric residency training programs generally do not require experience in ECT, and ECT is not tested in board certification examinations. Although the American Psychiatric Association offered guidelines for ECT practice and training in 1978, 1990, and 2002, few formal training programs exist. Training fellowships with hands-on experience are offered only at several academic medical centers in the United States. The number of their graduates is small.[20]

REPETITIVE TRANSCRANIAL MAGNETIC STIMULATION

rTMS, a tool initially intended to trace neural conduction pathways, was brought into clinical practice with the hope of emulating the remission rates achieved by ECT while simultaneously minimizing side effects. In rTMS, a magnetic coil is placed in close proximity to the scalp. When it is activated, a rapidly alternating current passes through the coil, inducing a magnetic field. The magnetic field then precipitates an electrical field in the targeted brain area, which results in neuronal depolarization.[21]

A number of variables can be altered to influence the outcome of rTMS, including the pulse frequency, the stimulation intensity, the number of pulses per treatment, the time between consecutive pulses, the sequence of pulses (number and intensity) across sessions, and the location at which the stimulus is applied. Still, no adjustments allow for stimulation of brain regions deeper than 1 cm below the skull.[22]

Regardless of treatment parameters, rTMS is generally well tolerated (see Loo *et al.*[23] for a comprehensive review). A transient mild headache is the most common side effect. Although there is a risk of seizure, induction of a manic state, and hearing impairment, the incidence of these adverse events is low and can be minimized with careful screening of potential patients and other simple precautions, such as earplugs. The data on neuropsychological impact are inconsistent with respect to specific changes precipitated by rTMS treatment. However, on the whole, loss of cognitive function after rTMS is rare.[23]

A number of blinded, sham-controlled studies evaluating the efficacy of rTMS have found that high-frequency rTMS of the left dorsolateral prefrontal cortex[24,25] or low-frequency rTMS of the right prefrontal cortex produces a significant reduction in depression severity.[26,27] Two recent investigations have documented the efficacy of bilateral rTMS, employing both high-frequency rTMS of the left dorsolateral prefrontal cortex and the right prefrontal cortex.[28,29] Indeed, several meta-analyses conducted to gauge the efficacy of rTMS in treating depression have concluded that rTMS is consistently superior to sham transcranial magnetic stimulation (TMS) in treating depression.[30]

However, the clinical improvement that participants experienced, although statistically significant in the trials examined, was generally judged to be modest and rarely durable.[30] For instance, Holtzheimer *et al.*,[31] combining the results of 12 studies, reported that rTMS reduced participants' Hamilton Rating Scale for Depression (HAM-D) scores by an average of 37% in open studies, 23.8% in active control studies, and 7.3% in sham-controlled studies. They further noted that the effect sizes in rTMS trials tend to be significantly smaller than those in ECT research.[31] This conclusion is supported by the data emerging from studies comparing ECT to rTMS treatment.[32–34]

The trends that emerged from the meta-analyses are congruent with the results of the pivotal Neuronetics-funded multicenter study designed to gain US Food and Drug Administration (FDA) regulatory approval in the United States; that study failed to show a significant difference between patients randomized to rTMS ($n = 155$) and sham ($n = 146$) when reductions in primary outcome measure scores [the Montgomery-Asberg Depression Rating Scale (MADRS)] were compared (mean MADRS reduction of -3.9 for sham rTMS versus -6 for active rTMS at 6 weeks, $P = 0.058$).[35] Study investigators contended that the statistical significance for MADRS outcome was just barely missed and that the superiority of active rTMS over sham rTMS was assessed by numerous secondary outcomes, including a mean reduction of -5.4 for active rTMS versus -3.3 for sham in the HAM-D17 score at week 6 ($P = 0.005$).[36] However, in January 2007, an FDA advisory panel did not agree, concluding that "the device's effectiveness was 'small,' 'borderline,' 'marginal' and 'of questionable clinical significance.'"[36]

Although this was certainly a setback for rTMS in the United States, the FDA has yet to issue a final decision regarding its approval. ClinicalTrials.gov presently lists 10 active studies evaluating the efficacy of the treatment in adults with unipolar depression, including an additional Neuronetics-funded multisite, randomized, controlled trial and 2 trials that specifically target TRD.[37]

HYBRID AND DERIVATIVE THERAPIES OF ECT AND RTMS

MST uses the same device as TMS but, similar to ECT, induces a therapeutic seizure. The purpose of using the TMS technology is to promote localization and decrease the intensity of the seizures in order to facilitate cognitive sparing. Only 2 published studies to date have investigated MST in patients with TRD. Employing a within-subject crossover design, Lisanby *et al.*,[38] preliminarily finding that MST elicited substantial mood improvement with side effects less severe than those usually observed with a similar course of ECT, showed the feasibility and safety of this approach. A follow-up study by White *et al.*[39] also demonstrated that MST, though effective in reducing depression severity, was less robust than a traditional course of ECT. Several ongoing clinical trials continue to evaluate MST's clinical potential (see www.clinicaltrials.gov).

Another hybrid of TMS and ECT, focal electrically administered stimulated seizure therapy, uses ECT devices but attempts to circumscribe the seizure to the prefrontal cortex. Here too, the goal is to reduce involvement of the medial temporal lobe in order to minimize the memory impairment associated with ECT.[16]

Transcranial direct current stimulation uses scalp electrodes to stimulate the brain via direct low-amplitude currents. The only published double-blind trial on this method to date found the stimulation of the left dorsolateral prefrontal cortex to be superior to that of occipital or sham transcranial direct current stimulation in reducing depression severity as measured by a statistically significant reduction in the HAM-D.[40] The low cost and noninvasive nature of this treatment make this an attractive candidate for further investigation. There are currently 3 ongoing clinical trials involving transcranial direct current stimulation for depression in the United States.[37]

Finally, chronic epidural cortical stimulation is a new method currently enrolling patients for its first assessment of benefit in depression. Epidural cortical stimulation involves cortical activation via a neurosurgically placed wire on the surface of the brain.[37]

VAGUS NERVE STIMULATION

VNS entails the implantation of a coil around the left cervical vagus nerve, which is then coupled to a pacemaker-like device surgically implanted in the left chest wall. The standard schedule of stimulation is 5 seconds every 30 minutes, 24 hours per day, although a number of parameters can be altered according to the patient's needs, including the signal frequency, on-time, pulse width, and output current.[41] It is considered relatively safe; common adverse events include hoarseness and other throat irritation, paresthesia, and dyspnea.[42] Most adverse events resolve over the first year of treatment.

VNS was developed by Cyberonics for the treatment of epilepsy, and its antidepressant effect was unexpected. During trials evaluating the efficacy of VNS as an intervention for treatment-resistant epilepsy, an elevation in mood was observed among some participants.[43] Although the epileptic patients did not suffer from major depressive disorder and the initial analyses of mood change did not show a statistically significant difference in outcome between the VNS and control groups, Cyberonics commissioned studies of VNS in TRD.

Post hoc theories of the efficacy of VNS hinge on the vagus nerve's efferent connections to the locus ceruleus, raphe nuclei, and nucleus tractus solitarius,[44] brain regions associated with norepinephrine and serotonin regulation. Neuroimaging studies have demonstrated that VNS alters cerebral regional blood flow similarly to traditional antidepressants.[45] Two open-label[46,47] and 1 randomized, sham-controlled trial[48] have been conducted to evaluate the efficacy of VNS in TRD. In the pilot study ($n = 60$), 30.5% of participants responded (50% decrease in the HAM-D28 score) to VNS, and 15.3% remitted (HAM-D28 score <11) after 12 weeks of VNS.[47] After 24 months, an additional 13.5% (44% total) achieved response, and 6.7% (22% total) achieved remission. The subsequent pivotal sham-controlled study (D02, $n = 237$) failed to detect a significant difference between participants' response to sham versus active stimulation at 12 weeks post-implantation on the primary outcome measure, the HRSD-24 (15.2% versus 10.0%, $P = 0.251$).[49] As a result, Cyberonics revised the parameters of the pivotal analyses. The D02 patients were followed naturalistically for 12 months. They continued receiving VNS but were also permitted to have adjunctive treatment as usual. A second cluster-matched cohort of TRD patients, receiving only treatment as usual (D04), was added for comparison. After 1 year, the D02 response rate was 30%; the D04 response rate was 13%.[50,51] The remission rates, also significantly different, were 17% for D02 and 7% for D04. For D02 participants, the maximum benefit was not observed until 12 to 24 months, at which point one-third had responded.[52] On the basis of these data, VNS received FDA approval for TRD in 2005.

In 2007, the Center for Medicaid and Medicare Services chose not to cover the cost of VNS; a subsequent appeal by Cyberonics of this decision was revoked. The review panel stated that "the evidence supporting efficacy of VNS is not strong." The approval of VNS by the FDA was based on the safety of this therapy, rather than its efficacy.[53]

DEEP BRAIN STIMULATION

DBS is arguably the first psychiatric treatment that is directly informed by our evolving understanding of the neural circuitry involved in depression. Although we are far from a complete understanding of the complex circuitry in this disorder, much progress has been made since the initial monoamine theory of depression. It is now generally believed that depression has a complex genetic, developmental, and environmental etiology, and its manifestations lie at a systems level within the corticostriatothalamic circuitry.[54,55] Neuroimaging studies have elucidated a large number of brain areas that are abnormal in patients, some that are hyperactive and some that are hypoactive. DBS is a technique aimed specifically at restoring the activity in some of these abnormally functioning areas.

History of DBS

The history of DBS dates back to the 1950s, when it was sporadically used in the treatment of chronic pain.[56] Its modern use, however, is credited to Benabid *et al.*,[57] who in the 1980s introduced DBS as an effective treatment for essential tremor and Parkinson's disease (PD). Since that time, over 30,000 patients have been implanted with DBS electrodes,[58] mostly for the treatment of movement disorders. The FDA approved DBS at the thalamus for the treatment of essential and Parkinsonian tremor in 1997 and stimulation at the subthalamic nucleus (STN) and internal globus pallidus for PD in 2002; a humanitarian device exemption was granted for the treatment of dystonia in 2003.[58] DBS is currently at various stages of investigation for its use in pain, epilepsy, Tourette syndrome, obsessive-compulsive disorder (OCD), and, most recently, major depression.

Technical Overview

The DBS procedure involves the implantation of bilateral stimulation electrodes into a predetermined brain region. The region of interest is first localized by stereotactic neuroimaging, and burr holes are drilled into the skull under local anesthesia. The use of a recording electrode prior to the placement of the stimulation electrode further ensures the precise localization of the anatomical target.

The only DBS leads currently approved by the FDA are manufactured by Medtronic, Inc.; they are about 1.3 mm in diameter and contain 4 contacts at the distal end, spanning either 7.5 (model 3389) or 10.5 mm (model 3387) of brain tissue. Each contact can be activated independently at the physician's discretion. Medtronic also produces a larger lead that is not yet approved by the FDA but has been used in some trials of DBS for OCD and major depression. Advanced Neuromodulation Systems, Inc., produces a DBS system that is currently used in several clinical trials but has not yet gained FDA approval.

Approximately 1 week after the stereotactic surgery, the electrodes are connected to a pulse generator implanted under the clavicle via a subcutaneously tunneled extension cable. The stimulation parameters are thereafter adjusted telemetrically. Stimulation is usually continuous, although tremor patients may turn their device(s) off at night. The stimulation parameters, which have been determined empirically and vary extensively, depending on the diagnosis and brain target, are voltage (1–10.5 V), pulse width (30–450 microseconds), and frequency (60–280 Hz). Optimal settings may be easily determined in some disorders such as essential tremor, in which tremor represents an objective, measurable symptom, and clinical benefit is obtained on the order of seconds. The programming process is much more challenging in psychiatric disorders, for which clinicians must rely on subjective symptoms, and a considerable lag time exists (on the order of months) between the induction of stimulation and clinical benefit.

Risks Associated with DBS

Generally, DBS is considered to be safe with low rates of complications. There are, however, some serious risks that need to be addressed when this treatment option is discussed with patients, including intracranial bleeding, seizure, infection, and hardware malfunction or breakage.[56] Stimulation-induced side effects such as paresthesia, muscle contraction, dysarthria, and diplopia are usually either temporary or reversible with stimulation parameter adjustment.[59] Although DBS is considered safe as a neurosurgical method, is should be noted that it also constitutes the riskiest and most invasive treatment for depression currently available. In addition, to date, only a handful of patients have received DBS specifically for depression, and because neurostimulation for other disorders such as PD involves distinctly different brain regions, it is difficult to ascertain a precise figure for the associated risks.

Use of DBS in Major Depressive Disorder

The use of DBS in depression is currently under investigation in the United States, with ongoing trials in subgenual cingulate Brodmann area 25 (BA25) and in the ventral anterior capsule/ventral striatum.[37] The first evidence that DBS can alter mood came from reports of side effects in DBS for movement disorders. Reports of depression following STN-DBS range from 1.5% to 25%.[60] Other behavioral and personality changes such as anxiety, hypersexuality, apathy, and aggressiveness have also been reported,[61] and suicidality can be as high as 4.3% after DBS for movement disorders, even in patients with no previous suicidal ideation and with excellent motor outcome.[62] Importantly, there have also been reports of improvement in depression and/or hypomania in PD-DBS,[60,61] even though the target was the same as that in patients who experienced deterioration in mood. In 1 review, 12 of 72 (20%) PD patients had an improvement in mood, whereas only 3 (3%) experienced a worsening of mood.[63] In addition, early DBS work targeting the brain stem and mesodiencephalon for the treatment of chronic pain showed promise in the treatment of mood disorders, probably via effects on catecholaminergic and serotonergic regulatory cells.[64] Eight of 10 patients showed an improvement in the Minnesota Multiphasic Personality Inventory T score for depression, with 3 of them experiencing full recovery.

One of the 2 most promising targets for DBS for depression is a region containing the anterior limb of the internal capsule and portions of the ventral striatum, including the nucleus accumbens. Although the specific location of the target has varied, this region has shown great promise in the treatment of highly treatment-resistant OCD in studies from France, Belgium, and the United States (see the article by Berlin and Hollander in this issue). In a classic case of scientific serendipity, it was found that OCD patients with comorbid depression experience improvement in both disorders after DBS. The improvement in depression was first reported in a handful of case reports, 2 of which targeted the internal capsule[65,66] and 1 of which targeted the ventral caudate nucleus.[67] Interestingly, the improvements in OCD and depression symptoms exhibited different time courses, with mood symptoms improving weeks to months before improvement in OCD. This is consistent with the observation that the pharmacological treatment of these 2 disorders also shows different kinetics, with improvement in depression within weeks and improvement in OCD within months, and argues for DBS having a real impact on depression circuitry rather than improving mood simply by improving OCD symptoms.

In the largest published examination of long-term outcome data on DBS for OCD, Greenberg *et al.*,[68] in a collaborative study, implanted electrodes in 10 OCD patients, 8 of whom had comorbid major depression. In that study, the electrodes were placed within the ventral half of the anterior limb of the internal capsule, with the distal-most contact extending into the ventral striatum, in the caudal nucleus accumbens; the authors referred to this region as the ventral capsule/ventral striatum (VC/VS) and noted that it was more posterior than the one used in other reports (eg, see Aouizerate *et al.*[65] and Abelson *et al.*[67]). The baseline HAM-D24 mean was 21.1 ± 1.5 for this cohort, and it decreased to 14.7 ± 2.1 by 3 months, thereafter remaining stable, with a mean of 15.4 ± 2.6 at 36 months. Adverse effects included a small asymptomatic intracerebral hemorrhage, a single intraoperative generalized tonic-clonic seizure (with no further recurrence during the 3-year follow-up period), a superficial surgical wound infection in a diabetic patient, and various stimulation-dependent changes in mood and motor function that were reversible within seconds to minutes of stimulation cessation or parametric changes. In a follow-up study, this group used this target in the treatment of TRD. To date, electrodes have been implanted in 15 patients, who have been followed for a minimum of 1 year (D. Malone, unpublished data, 2007). These patients have been treated with a mean of 10 different antidepressant medications and have had a mean of 33 ECT treatments. For this group of highly resistant patients, the baseline MADRS mean was 34.8, with a decrease to 18.5 at 1 year. At the last follow-up (1 year), 9 of 15 patients showed a greater than 50% reduction in MADRS, and 6 of these 9 were in remission (MADRS <10).

Most recently, a group from Germany also began targeting a region similar to the VC/VS, but with the main focus of stimulation within the nucleus accumbens.[69] This initial report on 3 patients (2 of whom were monozygotic twins) showed promise in treating the specific symptom of anhedonia; this is not surprising, given the role of this brain region in reward processing. Within 60 seconds of blinded stimulation onset, patients were noted to spontaneously report the desire for a new pleasurable activity without reporting an improvement in depression. For example, 1 patient informed the medical group that he realized he was in Cologne and that, having never visited the famous Cologne Cathedral, he would like to do so while there. When asked about depressive symptomatology, however, the patient denied any changes. Both HAM-D24 and MADRS scores decreased significantly within a week of stimulation onset and were significantly correlated with blinded on/off stimulation status and stimulation strength.

The other investigated target in DBS for major depression has been the subgenual cingulate region (BA25 or Cg25).[70] Mayberg *et al.*[70] chose this region as a target on the basis of neuroimaging studies that showed activity decreases in this area with alleviation of depressive symptoms by pharmacotherapy,[71] rTMS,[72] and ECT[73] as well as activity increases in this area following stimulus-induced sadness.[71,74] In addition, the TRD patients enrolled in Mayberg *et al.*'s study showed marked hyperactivity in BA25 not previously described. (Recently, BA25 overactivity has also been shown to predict future depression in 2 populations of cancer patients.[75,76]) Of the 6 patients included in the initial report, 4 were responders at 6 months (>50% reduction in the HAM-D17 score), and 2 of the 4 responders were in remission (HAM-D17 score <8). Adverse effects included 2 patients experiencing local infections to the scalp and chest, respectively, and 1 patient experiencing skin erosion over the hardware; at high stimulation settings, patients experienced lightheadedness and psychomotor slowing, which was immediately reversible upon parametric changes. The same research group recently expanded these data to 15 patients followed for at least 6 months: of these, 8 have achieved response, and 5 have achieved remission (H. Mayberg, unpublished data, 2007).

Additional targets have also been proposed for DBS in TRD. In a case report from a Mexican group, remission of major depression disorder was achieved with DBS delivered to the inferior thalamic peduncle, a system of fibers connecting midline and intralaminar thalamic nuclei of the nonspecific thalamic system with the orbitofrontal cortex (OFC).[77] Proposed but not yet tested targets include the rostral cingulate gyrus[78] and the lateral habenula.[79]

Mechanism of Action

DBS was first introduced into clinical practice because of the observation that high-frequency stimulation of the ventral intermediate nucleus of the thalamus elicited an effect of immediate tremor suppression identical to that observed with ablation of the same target. Because high-frequency stimulation was required to achieve this effect, it was originally hypothesized that DBS works by suppressing activity via depolarization blockade at the target nucleus. Decades of ensuing research into the mechanisms of action of DBS have not yet led to any definitive conclusions and at times have resulted in mutually exclusive hypotheses.

Before a discussion of possible mechanisms, it is important to outline some of the basic properties of electrical stimulation of brain matter. Multiple elements have the potential to be affected during stimulation: (1) the cell bodies of neurons surrounding the electrode, (2) axons originating in the stimulated nucleus whose cell bodies are too far from the electrode (efferent axons), (3) axons from other parts of the brain synapsing on the stimulated nucleus (afferent axons), and (4) axons both originating and terminating in different nuclei but passing close to the electrode (fibers of passage). In addition, the effect of stimulation on each cell type varies according to the intrinsic properties of the cell, such as channel composition and distribution. Thus, for example, stimulation in a rat thalamic slice showed 2 different types of cells: about two-thirds quickly depolarized but then repolarized and did not fire, whereas one-third depolarized and remained depolarized with variable firing.[80] Furthermore, a neuron's response varies with stimulation parameters such as the frequency, amplitude, pulse width, and duration. For example, in a study of rat STN slice stimulation, some (but not all) cells showed a dependence on frequency such that they remained actively firing during stimulation at 90 Hz but were completely suppressed at a frequency of 130 Hz.[81] Although we do not have a complete understanding of how stimulation affects each cell or cell type, some principles of excitability have been outlined: axons are more excitable than cell bodies, large myelinated axons are excited before small unmyelinated axons, fibers traveling parallel to the electrode are excited before fibers traveling transversely, and elements near the cathode are excited before elements near the anode. Furthermore, stimulation current does not spread as a simple cylinder around the electrode because of the differences in conductivity of various cell types; thus, 2 cells that are at an equal distance from the electrode can be affected differently according to what lies between each cell and the electrode.[82,83]

Investigations into the mechanisms of DBS have mainly focused on the circuitry involved in PD, and so caution must be used when these findings are extrapolated to the effects of DBS on depression circuitry. Extensive studies in both animal models and humans via neural recordings, neurochemical and neurogenomic approaches, and functional imaging have given rise to 4 main theories for the effects of DBS:

1. Depolarization blockade: by altering the activation of voltage-gated channels, DBS renders neurons incapable of generating action potentials.[84]
2. Synaptic inhibition: DBS preferentially activates inhibitory afferents that synapse on the targeted nucleus, therefore indirectly inhibiting the cells close to the stimulation site.[85]
3. Synaptic depression: cells at the stimulation site are excited, but because of the high frequency of stimulation, they quickly become depleted of neurotransmitters and fail to transmit signals.[86]
4. Jamming: stimulation modulates the pathological network activity.[87–89]

What should become immediately apparent is that the first 3 hypotheses predict that DBS works via inhibition, whereas the fourth hypothesis predicts that DBS works via excitation. These seemingly contradictory views have recently been linked by the neural modeling work of McIntyre *et al*.[88] Using multicompartment cable models of neurons coupled to extracellular electrical fields, this group was able to show 2 emergent properties of stimulation. First, activity in the cell body is not always correlated to or representative of spiking output generated in the axon. Thus, the efferent axon might be spiking, but an experiment that records from the cell body would not capture this activity. Second, close to the electrode, the stimulation is strong enough to excite cell bodies; however, around this sphere of excitation, the stimulation is strong enough only to excite axons; because the majority of afferent input into the STN is inhibitory, the net effect is a ring of inhibition surrounding the area of excitation. These modeling predictions are able to explain why in vivo recordings in the stimulated nucleus have shown decreased firing,[85,90–92] whereas recordings in a downstream nucleus have shown an increase in the output of the stimulated nucleus.[92–94]

In summary, the mechanism of DBS is not yet fully understood; however, it is thought to involve a complex mixture of inhibitory and excitatory actions. In PD, a key finding is a pathological rhythm in the STN of 11 to 30 Hz. It is currently believed that DBS stimulates STN cells, or at least their efferent axons, so that they follow the stimulation at 80 Hz or higher. This both prevents the low-frequency oscillation and replaces it with a "healthier" oscillation at a higher frequency.[89]

These results cannot be easily applied to the brain areas in TRD because the cellular content and properties in these regions are very different. To emphasize this point, it is interesting to note that the time course to clinical benefit differs dramatically for each of the disorders for which DBS has been used:

DBS improves tremors in seconds, gait disturbances and akinesia in minutes to hours, and dystonia in days to weeks,[82] whereas the clinical benefit in depression and OCD requires months.[68,70] Thus, it is possible that DBS neuromodulation in depression involves neural plasticity, that is, actual changes in synaptic connections rather than just alterations in firing rates.

Neuroimaging studies comparing preoperative and postoperative cerebral blood flow have demonstrated that DBS is effective in altering the circuitry involved in depression. At 3 months following the onset of DBS in BA25, there was a marked decrease in cerebral blood flow in this region and in the OFC, 2 areas with known involvement in the pathophysiology and treatment of depression.[70] In addition, responders to the treatment—but not nonresponders—also showed a decrease in metabolic activity in the hypothalamus, anterior insula, and medial frontal cortex and an increase in activity in the dorsolateral prefrontal cortex, dorsal anterior and posterior cingulate cortices, and premotor and parietal cortices. Chronic (3–26 months) stimulation in the VC/VS decreased activity in the subgenual anterior cingulate cortex (Brodmann area 32), the right left dorsolateral prefrontal cortex, and the right anterior insula.[95] Improvement in depression severity was inversely correlated with a decrease in activity in the left hippocampus and correlated positively with a decrease in activity in the right dorsolateral prefrontal cortex. Interestingly, activity in the ventral striatum, which was found to be hyperactive in these patients preoperatively and was a target of the stimulation, was not suppressed; in fact, a negative correlation was found between a lowered metabolic rate in the ventral striatum and an improvement in clinical scores. Thus, although DBS delivered to the subgenual cingulate cortex may work by inhibiting the target region, DBS to the VC/VS invokes a very different mechanism, perhaps one of altered activation patterns or neural jamming. Complicating the matter further is the finding that in acute stimulation to the VC/VS (ie, when DBS-on and DBS-off conditions are being compared), the OFC, primarily on the right side, is activated.[96] Acute stimulation also increases activation of the right striatum. The right-sided lateralization is particularly interesting and unexplained in light of the bilateral nature of DBS. In summary, the picture that emerges is the following: stimulation in the VC/VS directly activates the OFC and striatum, but over time, this activation changes the circuitry so as to cause the OFC to become less active than at baseline, whereas striatal activity remains roughly unchanged.

DBS: Summary

The initial reports of positive responses to DBS are important given the severe debilitation of the patient population under investigation. Although this procedure is in its infancy and only a handful of TRD patients have been treated, response rates of over 50% in patients that have tried dozens of medications and several trials of ECT are highly encouraging. There are, however, several drawbacks to DBS, most notably its prohibitive cost and the enormous amount of time and clinical expertise required both preoperatively and postoperatively. The long-term effects are unknown, and the potential complications, such as hemorrhage and infection, though occurring at very low rates, are severe enough to warrant careful consideration. In addition, ethical concerns have been raised by the notion of altering cognition and emotion by direct interference with the brain's neural activity.

FUTURE DIRECTIONS

Until recently, treatments for major depressive disorder lacked specificity because of large gaps in our understanding of the underlying etiology and pathogenesis. The advent of sophisticated neuroimaging technology has allowed us to begin unraveling the neural circuitry giving rise to this highly complex disorder involving a large number of brain regions and neurotransmitter systems.[97] Thus, we are finally entering an era in which scientific advances may directly inform clinical practice in psychiatry. DBS is the first treatment with the visionary goal of interfering with the pathological neural activity present in depression. As our understanding of neural circuitry advances, this method is likely to become more sophisticated. In epilepsy, for example, a neurofeedback device has been developed and is currently being investigated in patients. The closed-loop responsive neurostimulation system uses hardware similar to that used for DBS, but it is able to sense pathological neural firing, and stimulation pulses are given only as needed.[98] Responsive neurostimulation could one day allow for a permanently implantable yet not always active treatment option for TRD. Other potential advances in DBS include (1) more sophisticated stimulation paradigms that aim to better emulate neuronal activity, (2) multiple electrodes that target all of the regions with aberrant activity in an individual patient's brain, and (3) improvements in the hardware. One important area of research is the extension of battery life and the development of rechargeable batteries because in some instances in which large amounts of current are required to achieve clinical effect (eg, OCD), patients need to undergo surgical replacement of their pulse generators frequently. As basic science advances, new technologies are also likely to emerge. Recently, a method was developed for optical excitation and inhibition of neurons.[99] It is not hard to envision a future in which fiber optics replace the current wire leads, allowing for a more precise excitation focus with less damage to surrounding tissue.

Another important issue in the development of new techniques and treatments is the FDA regulatory process. The decision by the Center for Medicaid and Medicare Services not to cover the cost of VNS despite the treatment's approval by the FDA raises important questions about the FDA's approval process for new devices; the pivotal studies for both VNS and rTMS failed to show significance on the primary outcome, but VNS was approved whereas rTMS was not, even though many more studies evaluating the safety and efficacy of rTMS had been published at the time of its application. Although the FDA generally requires documentation of 2 adequate and well-controlled studies, it will accept as sufficient proof of effectiveness a single investigation when it is accompanied by other convincing data from uncontrolled research if a device or drug has already been approved for treatment of a closely related disease. Thus, the prior approval of VNS for epilepsy may have facilitated its approval for TRD. In addition, the FDA's guidelines for an "adequate and well-controlled study" are vague and wide open for interpretation. They include statements such as "the methods of assessment of subject's response are well-defined and reliable."[100]

Furthermore, relevant to all neurostimulatory devices is the fact that the FDA does not give recommendations regarding optimal stimulation parameters, in contrast to its recommended dosing guidelines for medications. This oversight parallels the inconsistency in treatment parameters used by practitioners and researchers and likely contributes to inadequate dosing regimens. Further study of optimal stimulation parameters is essential as more types of neurostimulatory devices are developed and considered for approval by the FDA.

In conclusion, TRD is undoubtedly a highly disabling disorder and often leaves patients desperate for new approaches. Safely administered and effective new treatments are necessary. DBS is the newest neuromodulatory advance in this field and may offer promise for these patients. However, DBS is likely not a final stop but rather one of many new targeted neurostimulatory approaches for treatment in psychiatry.

ACKNOWLEDGMENT

This research was supported by National Institute of Mental Health grants K23-MH-069656 and MO1-RR-0007 (to the General Clinical Research Center of the Mount Sinai School of Medicine) and by a National Alliance for Research on Schizophrenia and Depression Young Investigator Award (to Sanjay J. Mathew).

DISCLOSURES

There is nothing to report.

REFERENCES

1. Schildkraut JJ. The catecholamine hypothesis of affective disorders: a review of supporting evidence. *Am J Psychiatry* 1965; 122: 509–522.
2. Rush AJ, Trivedi M, Wisniewski S, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry* 2006; 163: 1905–1917.
3. Hollon SD, DeRubeis RJ, Shelton RC, et al. Prevention of relapse following cognitive therapy vs medications in moderate to severe depression. *Arch Gen Psychiatry* 2005; 62: 417–422.
4. Oquendo MA, Baca-Garcia E, Kartachov A, et al. A computer algorithm for calculating the adequacy of antidepressant treatment in unipolar and bipolar depression. *J Clin Psychiatry* 2003; 64: 825–833.
5. Prudic J, Haskett RF, Mulsant B, et al. Resistance to antidepressant medications and short-term clinical response to ECT. *Am J Psychiatry* 1997; 154: 721–722.
6. Prudic J, Sackeim HA, Devanand DP. Medication resistance and clinical response to electroconvulsive therapy. *Psychiatry Res* 1990; 31: 287–296.
7. Rush AJ, Thase M, Dubé S. Research issues in the study of difficult-to-treat depression. *Biol Psychiatry* 2003; 53: 743–753.
8. Thase ME, Rush AJ. When at first you don't succeed: sequential strategies for antidepressant nonresponders. *J Clin Psychiatry* 1997; 58(suppl 13): 23–29.
9. Fava M. Diagnosis and definition of treatment-resistant depression. *Biol Psychiatry* 2003; 53: 649–659.
10. Souery D, Amsterdam J, de Montigny C, et al. Treatment resistant depression: methodological overview and operational criteria. *Eur Neuropsychopharmacol* 1999; 9: 83–91.
11. Souery D, Oswald P, Massat I, et al. Clinical factors associated with treatment resistance in major depressive disorder: results from a European multicenter study. *J Clin Psychiatry* 2007; 68: 1062–1070.
12. Kessler RC, Merikangas KR, Wang PS. Prevalence, comorbidity, and service utilization for mood disorders in the United States at the beginning of

the twenty-first century. *Annu Rev Clin Psychol* 2007; 3: 137–158.

13. Lebensohn Z. The history of electroconvulsive therapy in the United States and its place in American psychiatry: a personal memoir. *Compr Psychiatry* 1999; 40: 173–181.
14. Kellner C, Knapp R, Petrides G, et al. Continuation electroconvulsive therapy vs pharmacotherapy for relapse prevention in major depression: a multisite study from the Consortium for Research in Electroconvulsive Therapy (CORE). *Arch Gen Psychiatry* 2006; 63: 1337–1344.
15. Burt C, McCaig L, Rechtseiner E. Ambulatory medical care utilization estimates for 2005. *Adv Data* 2007; 29: 1–15.
16. Mathew SJ, Amiel. Electroconvulsive therapy in treatment-resistant depression. *Prim Psychiatry* 2005; 12: 51–56.
17. Dowman J, Patel A, Rajput K. Electroconvulsive therapy: attitudes and misconceptions. *J ECT* 2005; 21: 84–87.
18. Sackeim H, Prudic J, Devanand DP, et al. A prospective, randomized, double-blind comparison of bilateral and right unilateral electroconvulsive therapy at different stimulus intensities. *Arch Gen Psychiatry* 2000; 57: 425–434.
19. Sackeim HA, Prudic J, Fuller R, et al. The cognitive effects of electroconvulsive therapy in community settings. *Neuropsychopharmacology* 2007; 32: 244–254.
20. Fink M, Taylor M. Electroconvulsive therapy: evidence and challenges. *JAMA* 2007; 298: 330–332.
21. George MS, Wassermann EM, Post RM. Transcranial magnetic stimulation: a neuropsychiatric tool for the 21st century. *J Neuropsychiatry Clin Neurosci* 1996; 8: 373–382.
22. Roth Y, Amir A, Levkovitz Y, Zangen A. Three-dimensional distribution of the electric field induced in the brain by transcranial magnetic stimulation using figure-8 and deep H-coils. *J Clin Neurophysiol* 2007; 24: 31–38.
23. Loo CK, McFarquhar TF, Mitchell PB. A review of the safety of repetitive transcranial magnetic stimulation as a clinical treatment for depression. *Int J Neuropsychopharmacol* 2008; 11: 131–147.
24. Avery D, Holtzheimer P III, Fawaz W, et al. A controlled study of repetitive transcranial magnetic stimulation in medication-resistant major depression. *Biol Psychiatry* 2006; 59: 187–194.
25. Berman R, Narasimhan M, Sanacora G, et al. A randomized clinical trial of repetitive transcranial magnetic stimulation in the treatment of major depression. *Biol Psychiatry* 2000; 47: 332–337.
26. Menkes DL, Bodnar P, Ballesteros RA, Swenson MR. Right frontal lobe slow frequency repetitive transcranial magnetic stimulation (SF r-TMS) is an effective treatment for depression: a case-control pilot study of safety and efficacy. *J Neurol Neurosurg Psychiatry* 1999; 67: 113–115.
27. Klein E, Kreinin I, Chistyakov A, et al. Therapeutic efficacy of right prefrontal slow repetitive transcranial magnetic stimulation in major depression: a double-blind controlled study. *Arch Gen Psychiatry* 1999; 56: 315–320.
28. Fitzgerald P, Benitez J, de Castella A, et al. A randomized, controlled trial of sequential bilateral repetitive transcranial magnetic stimulation for treatment-resistant depression. *Am J Psychiatry* 2006; 163: 88–94.
29. Stern W, Tormos J, Press D, et al. Antidepressant effects of high and low frequency repetitive transcranial magnetic stimulation to the dorsolateral prefrontal cortex: a double-blind, randomized, placebo-controlled trial. *J Neuropsychiatry Clin Neurosci* 2007; 19: 179–186.
30. Loo CK, Mitchell PB. A review of the efficacy of transcranial magnetic stimulation (TMS) treatment for depression, and current and future strategies to optimize efficacy. *J Affect Disord* 2005; 88: 255–267.
31. Holtzheimer PE, Russo J, Avery DH. A meta-analysis of repetitive transcranial magnetic stimulation in the treatment of depression. *Psychopharmacol Bull* 2001; 35: 149–169.
32. Grunhaus L, Shcriber S, Dolberg OT, et al. A randomized controlled comparison of electroconvulsive therapy and repetitive transcranial magnetic stimulation in severe and resistant nonpsychotic major depression. *Biol Psychiatry* 2003; 53: 324–331.
33. Janicak PG, Dowd SM, Martis B, et al. Repetitive transcranial magnetic stimulation versus electroconvulsive therapy for major depression: preliminary results of a randomized trial. *Biol Psychiatry* 2002; 51: 659–667.
34. Pridmore S, Bruno R, Turnier-Shea Y, et al. Comparison of unlimited numbers of rapid transcranial magnetic stimulation (rTMS) and ECT treatment sessions in major depressive episode. *Int J Neuropsychopharmacol* 2000; 3: 129–134.
35. O'Reardon J, Solvason HB, Janicak P, et al. Efficacy and safety of transcranial magnetic stimulation in the acute treatment of major depression: a multisite randomized controlled trial. *Biol Psychiatry* 2007; 62: 1208–1216.
36. U.S. Food and Drug Administration Center for Devices and Radiological Health. Neurological Devices Panel meeting on January 26, 2007. Available at: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfadvisory/details.cfm?mtg=675. Accessed May 2008.
37. ClinicalTrials.gov. Available at: http://www.clinicaltrials.gov. Accessed February 2008.
38. Lisanby SH, Luber B, Schlaepfer TE, Sackeim HA. Safety and feasibility of magnetic seizure therapy (MST) in major depression: randomized within-subject comparison with electroconvulsive therapy. *Neuropsychopharmacology* 2003; 28: 1852–1865.
39. White PF, Amos Q, Zhang Y, et al. Anesthetic considerations for magnetic seizure therapy: a novel therapy for severe depression. *Anesth Analg* 2006; 103: 76–80.
40. Boggio PS, Rigonatti SP, Ribeiro RB, et al. A randomized, double-blind clinical trial on the efficacy of cortical direct current stimulation for the treatment of major depression. *Int J Neuropsychopharmacol* 2008; 11: 249–254.
41. George MS, Sackeim HA, Marangell LB, et al. Vagus nerve stimulation. A potential therapy for resistant depression? *Psychiatr Clin North Am* 2000; 23: 757–783.
42. Uthman BM, Reichl AM, Dean JC, et al. Effectiveness of vagus nerve stimulation in epilepsy patients: a 12-year observation. *Neurology* 2004; 63: 1124–1126.

43. Harden CL, Pulver MC, Ravdin LD, et al. A pilot study of mood in epilepsy patients treated with vagus nerve stimulation. *Epilepsy Behav* 2000; 1: 93–99.
44. VanBockstaele E, Peoples J, Valentino R. Anatomic basis for differential regulation of the rostrolateral peri-locus coeruleus region by limbic afferents. *Biol Psychiatry* 1999; 46: 1352–1363.
45. Conway C, Sheline Y, Chibnall J, et al. Cerebral blood flow changes during vagus nerve stimulation for depression. *Psychiatry Res Neuroimaging* 2006; 146: 179–184.
46. Rush AJ, George M, Sackeim H, et al. Vagus nerve stimulation (VNS) for treatment-resistant depressions: a multicenter study. *Biol Psychiatry* 2000; 47: 276–286.
47. Sackeim HA, Rush AJ, George MS, et al. Vagus nerve stimulation (VNS) for treatment-resistant depression: efficacy, side effects, and predictors of outcome. *Neuropsychopharmacology* 2001; 25: 713–728.
48. Rush AJ, Marangell L, Sackeim H, et al. Vagus nerve stimulation for treatment-resistant depression: a randomized, controlled acute phase trial. *Biol Psychiatry* 2005; 58: 347–354.
49. Rush AJ, Sackeim H, Marangell L, et al. Effects of 12 months of vagus nerve stimulation in treatment-resistant depression: a naturalistic study. *Biol Psychiatry* 2005; 58: 355–363.
50. Marangell LB, Rush AJ, George MS, et al. Vagus nerve stimulation (VNS) for major depressive episodes: one year outcomes. *Biol Psychiatry* 2002; 51: 280–287.
51. George MS, Rush AJ, Marangell LB, et al. A one-year comparison of vagus nerve stimulation with treatment as usual for treatment-resistant depression. *Biol Psychiatry* 2005; 58: 364–373.
52. Nahas Z, Marangell LB, Husain MM, et al. Two-year outcome of vagus nerve stimulation (VNS) for treatment of major depressive episodes. *J Clin Psychiatry* 2005; 66: 1097–1104.
53. Phurrough S, Salive MS, Lofton B, Schafer J. Decision memo for vagus nerve stimulation for treatment of resistant depression (TRD) (CAG-00313R). Available at: https://www.cms.hhs.gov/mcd/viewdecisionmemo.asp?id=195. Accessed May 2008.
54. Rauch SL. Neuroimaging and neurocircuitry models pertaining to the neurosurgical treatment of psychiatric disorders. *Neurosurg Clin N Am* 2003; 14: 213–223, vii–viii.
55. Mayberg HS. Modulating dysfunctional limbic-cortical circuits in depression: towards development of brain-based algorithms for diagnosis and optimised treatment. *Br Med Bull* 2003; 65: 193–207.
56. Anderson WS, Lenz FA. Surgery insight: Deep brain stimulation for movement disorders. *Nat Clin Pract Neurol* 2006; 2: 310–320.
57. Benabid AL, Pollak P, Louveau A, et al. Combined (thalamotomy and stimulation) stereotactic surgery of the VIM thalamic nucleus for bilateral Parkinson disease. *Appl Neurophysiol* 1987; 50: 344–346.
58. Medtronic, Inc. Available at: http://www.medtronic.com. Accessed September 2007.
59. Eitan R, Lerer B. Nonpharmacological, somatic treatments of depression: electroconvulsive therapy and novel brain stimulation modalities. *Dialogues Clin Neurosci* 2006; 8: 241–258.
60. Voon V, Kubu C, Krack P, et al. Deep brain stimulation: neuropsychological and neuropsychiatric issues. *Mov Disord* 2006; 21(suppl 14): S305–S327.
61. Temel Y, Kessels A, Tan S, et al. Behavioural changes after bilateral subthalamic stimulation in advanced Parkinson disease: a systematic review. *Parkinsonism Relat Disord* 2006; 12: 265–272.
62. Burkhard PR, Vingerhoets FJ, Berney A, et al. Suicide after successful deep brain stimulation for movement disorders. *Neurology* 2004; 63: 2170–2172.
63. Castelli L, Perozzo P, Zibetti M, et al. Chronic deep brain stimulation of the subthalamic nucleus for Parkinson's disease: effects on cognition, mood, anxiety and personality traits. *Eur Neurol* 2006; 55: 136–144.
64. Andy OJ, Jurko F. Thalamic stimulation effects on reactive depression. *Appl Neurophysiol* 1987; 50: 324–329.
65. Abelson JL, Curtis GC, Sagher O, et al. Deep brain stimulation for refractory obsessive-compulsive disorder. *Biol Psychiatry* 2005; 57: 510–516.
66. Gabriels L, Cosyns P, Nuttin B, et al. Deep brain stimulation for treatment-refractory obsessive-compulsive disorder: psychopathological and neuropsychological outcome in three cases. *Acta Psychiatr Scand* 2003; 107: 275–282.
67. Aouizerate B, Cuny E, Martin-Guehl C, et al. Deep brain stimulation of the ventral caudate nucleus in the treatment of obsessive-compulsive disorder and major depression. Case report. *J Neurosurg* 2004; 101: 682–686.
68. Greenberg BD, Malone DA, Friehs GM, et al. Three-year outcomes in deep brain stimulation for highly resistant obsessive-compulsive disorder. *Neuropsychopharmacology* 2006; 31: 2384–2393.
69. Schlaepfer TE, Cohen MX, Frick C, et al. Deep brain stimulation to reward circuitry alleviates anhedonia in refractory major depression. *Neuropsychopharmacology* 2008; 33: 368–377.
70. Mayberg HS, Lozano AM, Voon V, et al. Deep brain stimulation for treatment-resistant depression. *Neuron* 2005; 45: 651–660.
71. Mayberg HS, Liotti M, Brannan SK, et al. Reciprocal limbic-cortical function and negative mood: converging PET findings in depression and normal sadness. *Am J Psychiatry* 1999; 156: 675–682.
72. Mottaghy FM, Keller CE, Gangitano M, et al. Correlation of cerebral blood flow and treatment effects of repetitive transcranial magnetic stimulation in depressed patients. *Psychiatry Res Neuroimaging* 2002; 115: 1–14.
73. Nobler MS, Oquendo MA, Kegeles LS, et al. Decreased regional brain metabolism after ECT. *Am J Psychiatry* 2001; 158: 305–308.
74. Damasio AR, Grabowski TJ, Bechara A, et al. Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nat Neurosci* 2000; 3: 1049–1056.
75. Inagaki M, Yoshikawa E, Kobayakawa M, et al. Regional cerebral glucose metabolism in patients with secondary depressive episodes after fatal pancreatic cancer diagnosis. *J Affect Disord* 2007; 99: 231–236.
76. Kumana H, Ida I, Oshima, et al. Brain metabolic changes associated with predisposition to onset of major depressive disorder in cancer patients–a

preliminary PET study. *J Psychiatr Res* 2007; 41: 591–599.

77. Jimenez F, Velasco F, Salin-Pascual R, et al. A patient with a resistant major depression disorder treated with deep brain stimulation in the inferior thalamic peduncle. *Neurosurgery* 2005; 57: 585–593; discussion 585–593.
78. Sakas DE, Panourias IG. Rostral cingulate gyrus: a putative target for deep brain stimulation in treatment-refractory depression. *Med Hypotheses* 2006; 66: 491–494.
79. Sartorius A, Henn FA. Deep brain stimulation of the lateral habenula in treatment resistant major depression. *Med Hypotheses* 2007; 69: 1305–1308.
80. Anderson T, Hu B, Pittman Q, Kiss ZH. Mechanisms of deep brain stimulation: an intracellular study in rat thalamus. *J Physiol* 2004; 559: 301–313.
81. Magarinos-Ascone C, Pazo JH, Macadar O, Buno W. High-frequency stimulation of the subthalamic nucleus silences subthalamic neurons: a possible cellular mechanism in Parkinson's disease. *Neuroscience* 2002; 115: 1109–1117.
82. Lozano AM, Eltahawy H. How does DBS work? *Suppl Clin Neurophysiol* 2004; 57: 733–736.
83. Perlmutter JS, Mink JW. Deep brain stimulation. *Annu Rev Neurosci* 2006; 29: 229–257.
84. Beurrier C, Bioulac B, Audin J, Hammond C. High-frequency stimulation produces a transient blockade of voltage-gated currents in subthalamic neurons. *J Neurophysiol* 2001; 85: 1351–1356.
85. Dostrovsky JO, Levy R, Wu JP, et al. Microstimulation-induced inhibition of neuronal firing in human globus pallidus. *J Neurophysiol* 2000; 84: 570–574.
86. Urbano FJ, Leznik E, Llinas RR. Cortical activation patterns evoked by afferent axon stimuli at different frequencies: an in vitro voltage-sensitive dye imaging study. *Thalamus Rel Syst* 2002; 1: 371–378.
87. Montgomery EB Jr, Baker KB. Mechanisms of deep brain stimulation and future technical developments. *Neurol Res* 2000; 22: 259–266.
88. McIntyre CC, Savasta M, Kerkerian-Le Goff L, Vitek JL. Uncovering the mechanism(s) of action of deep brain stimulation: activation, inhibition, or both. *Clin Neurophysiol* 2004; 115: 1239–1248.
89. Garcia L, D'Alessandro G, Bioulac B, Hammond C. High-frequency stimulation in Parkinson's disease: more or less? *Trends Neurosci* 2005; 28: 209–216.
90. Benazzouz A, Piallat B, Pollak P, Benabid AL. Responses of substantia nigra pars reticulata and globus pallidus complex to high frequency stimulation of the subthalamic nucleus in rats: electrophysiological data. *Neurosci Lett* 1995; 189: 77–80.
91. Boraud T, Bezard E, Bioulac B, Gross C. High frequency stimulation of the internal globus pallidus (GPi) simultaneously improves Parkinsonian symptoms and reduces the firing frequency of GPi neurons in the MPTP-treated monkey. *Neurosci Lett* 1996; 215: 17–20.
92. Maurice N, Thierry AM, Glowinski J, Deniau JM. Spontaneous and evoked activity of substantia nigra pars reticulata neurons during high-frequency stimulation of the subthalamic nucleus. *J Neurosci* 2003; 23: 9929–9936.
93. Hashimoto T, Elder CM, Okun MS, et al. Stimulation of the subthalamic nucleus changes the firing pattern of pallidal neurons. *J Neurosci* 2003; 23: 1916–1923.
94. Anderson ME, Postupna N, Ruffo M. Effects of high-frequency stimulation in the internal globus pallidus on the activity of thalamic neurons in the awake monkey. *J Neurophysiol* 2003; 89: 1150–1160.
95. Van Laere K, Nuttin B, Gabriels L, et al. Metabolic imaging of anterior capsular stimulation in refractory obsessive-compulsive disorder: a key role for the subgenual anterior cingulate and ventral striatum. *J Nucl Med* 2006; 47: 740–747.
96. Rauch SL, Dougherty DD, Malone D, et al. A functional neuroimaging investigation of deep brain stimulation in patients with obsessive-compulsive disorder. *J Neurosurg* 2006; 104: 558–565.
97. Ressler KJ, Mayberg HS. Targeting abnormal neural circuits in mood and anxiety disorders: from the laboratory to the clinic. *Nat Neurosci* 2007; 10: 1116–1124.
98. Fountas KN, Smith JR. A novel closed-loop stimulation system in the control of focal, medically refractory epilepsy. *Acta Neurochir Suppl* 2007; 97: 357–362.
99. Zhang F, Wang LP, Brauner M, et al. Multimodal fast optical interrogation of neural circuitry. *Nature* 2007; 446: 633–639.
100. FDA guidelines. Available at: http://www.fda.gov. Accessed February 2008.

Treating Pediatric Patients with Antipsychotic Drugs: Balancing Benefits and Safety

Iliyan Ivanov, MD, and Alexander Charney, BA

Department of Psychiatry, Mount Sinai School of Medicine, New York, NY

In the last 2 decades, the advances of evidence-based medicine have had a palpable influence on psychiatry. In pediatric psychiatry, this has led to increased acceptance of the idea that neuropsychiatric conditions of childhood are in part biologically determined[1,2] and that the use of psychotropic agents can remediate some of the symptoms.[3] The use of various psychotropic agents for children suffering from these conditions has become increasingly widespread and accepted by both medical professionals and society at large. Like any other biological treatment, the various types of psychotropic drugs present a number of adverse effects, ranging from mild to severe and potentially life-threatening. Among psychotropic compounds, antipsychotic agents in particular have received considerable attention, which is in part due to their efficacy and in part due to the spectrum of serious side effects that they produce.[4] This article reviews the current trends for the therapeutic use of antipsychotic medications in pediatric psychiatry and the available data that support their efficacy, but it also discusses important issues related to their safety during both short-term and long-term administration.

ANTIPSYCHOTIC/NEUROLEPTIC AGENTS

The terms *antipsychotics* and *neuroleptics* refer to several classes of pharmacological agents possessing a common antagonistic effect on the brain dopamine receptors that enables them to alleviate psychotic symptoms (ie, hallucinations, delusions, and disorganized thinking and behavior). In addition to a strictly pharmacological classification that includes phenothiazine, piperazine, piperidine, thioxanthene, butyrophenone, and heterocyclic classes, the antipsychotic agents are broadly divided into typical and atypical antipsychotics (for details, see Table 1). Initially developed in the 1950s, typical antipsychotics are agents that have a high affinity for the D2-dopamine receptors and the potential to cause severe neuromotor disturbances (ie, iatrogenic Parkinsonism, akathisia, dystonia, and tardive dyskinesia). Atypical antipsychotics are novel agents (Table 1) that presumably cause fewer acute and chronic neuromotor effects.[5] The severity of side effects varies among agents, for each atypical agent exhibits its own unique affinity to the D2 receptors. Several of these new agents were introduced and approved for the treatment of psychotic disorders in adults during the 1990s, and this led to an increase in their use among both adults and younger patients. As the prevalence of the use of atypical agents rises, new evidence is emerging that suggests that the previous notion of a safer side effect profile associated with these antipsychotics may need to be re-examined. Because side effects such as hyperprolactinemia, weight gain, and the development of metabolic syndrome and diabetes may have long-term consequences, the use of these agents in children and adolescents should be carefully considered, and their clinical efficacy should be balanced against issues of safety.

ANTIPSYCHOTIC DRUGS APPROVED FOR USE IN CHILDREN

A few of the typical antipsychotic agents have been approved by the Food and Drug Administration

Address Correspondence to:

Iliyan Ivanov
Department of Psychiatry
Mount Sinai School of Medicine
New York, NY
Email: iliyan.ivanov@mssm.edu

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20051

© 2008 Mount Sinai School of Medicine

Table 1. *Classes of Antipsychotic Agents.*

Class	Agent	FDA Approval for Use in Adults	FDA Approval for Use in Children and Adolescents	FDA Daily Dose Range for Children and Adolescents
Typical antipsychotics				
High-potency	Haloperidol	Schizophrenia	Psychotic d/o, nonpsychotic behavior d/o, and Tourette's d/o in children ages 3 to 12	0.05 to 0.15 mg/kg
	Trifluoperazine	Schizophrenia and nonpsychotic anxiety d/o	Psychotic d/o in children ages 6 to 12	1 to 15 mg
	Perphenazine	Schizophrenia and nausea/emesis	Not approved for children <12 years of age	5 to 10 mg
	Thiothixene	Schizophrenia	Not approved for children <12 years of age	6 to 60 mg
	Fluphenazine	Schizophrenia	Not approved for children and adolescents	N/A
	Pimozide	Tourette's d/o	Tourette's d/o in children >12 years of age	0.05 to 0.2 mg/kg
Mid-potency	Molindone	Schizophrenia	Not approved for children <12 years of age	50 to 225 mg
	Loxepine	Schizophrenia	Not approved for children and adolescents	N/A
Low-potency	Thioridazine	Schizophrenia	Psychotic and behavior d/o in children ages 2 to 12	0.5 to 3 mg/kg
	Mezoridazine	Schizophrenia and behavioral problems with chronic brain syndrome	Not approved for children and adolescents	N/A
	Chlorpromazine	Schizophrenia, nausea/emesis, intermittent porphyria, intractable hiccups, bipolar d/o, and tetanus	Behavioral problems and psychotic d/o in children ages 6 months to 12 years	40 mg for 6 to 60 months and 75 mg for 5 to 12 years
Atypical antipsychotics	Clozapine	Schizophrenia	Not approved for children and adolescents	N/A
	Risperidone	Schizophrenia	Schizophrenia for ages 13 to 17, bipolar d/o for ages 10 to 17, and agitation with autism for ages 5 to 17	1 to 6 mg; 0.25 mg for patients <20 kg and 0.5 mg for patients ≥20 kg
	Olanzapine	Schizophrenia and bipolar d/o	Not approved for children and adolescents	N/A
	Quetiapine	Schizophrenia and bipolar d/o	Not approved for children and adolescents	N/A
	Ziprasidone	Schizophrenia and bipolar d/o	Not approved for children and adolescents	N/A
	Aripiprazole	Schizophrenia, bipolar d/o, and add-on treatment for MDD	Schizophrenia for ages 13 to 17	2 to 30 mg
	Paliperidone	Schizophrenia	Not approved for children and adolescents	N/A
	Asenapine Iloperidone	FDA has accepted new drug applications for both agents as of November 2007.	N/A	N/A

Abbreviations: d/o, disorder; FDA, Food and Drug Administration; MDD, major depressive disorder; N/A, not applicable.

(FDA) for use in the treatment of childhood psychotic and behavior disorders (see Table 1). Haloperidol, trifluoperazine, and chlorpromazine have been approved for the treatment of psychosis, and pimozide has been approved for the treatment of Tourette's syndrome in children 12 years of age and younger. Although several other typical agents have been approved for the treatment of schizophrenia/psychotic disorders in adults, the FDA labeling on these medications specifies that they are not recommended for use in children younger than 12 (although they could be used in adolescents at minimal adult doses; see Table 1).

Recently, 2 atypical antipsychotics received approval by the FDA for their use in children and adolescents. In 2006, risperidone was approved for the treatment of irritability associated with autism in children ages 5 to 16, and a year later, it was approved for both the short-term treatment of schizophrenia-related symptoms in adolescents ages 13 to 17 and the short-term treatment of bipolar disorder in children and adolescents ages 10 to 17. The second agent is aripiprazole, which was initially approved by the FDA in 2007 for the short-term treatment of schizophrenia in adolescents 13 to 17 years of age and as recently as March 2008 received a FDA approval for the short-term treatment of bipolar disorder in adolescents ages 10 to 17. The majority of trials that provided data to support these approvals were conducted by the drugs' manufacturers and demonstrated that the efficacy of both drugs was superior to that of placebo.

Although it is true that the formal approval of a particular agent for the treatment of a particular disorder tends to stimulate the prescribing practices of health professionals, concerns about the overprescribing of antipsychotics for children and adolescents were raised long before the recent FDA approvals. In the following section, we review data from epidemiologic studies that may inform both clinicians and the general public about trends in the use of antipsychotic drugs in children over the last 2 decades.

PREVALENCE OF ANTIPSYCHOTIC USE IN CHILDREN AND ADOLESCENTS

There are a number of sources that provide information related to the use of antipsychotics in young patients. Evidence from studies published in peer-reviewed journals supports the notion that the use of antipsychotics in children has increased significantly in recent years. One study found that the number of children ages 2 to 18 in the Tennessee Medicaid program who were prescribed antipsychotic drugs nearly doubled from 22.9 per 10,000 children in 1996 to 45.4 per 10,000 children in 2001.[6] Another study reviewed 3 Medicaid state programs and 1 private insurance organization and found that total antipsychotic use increased approximately 3-fold from 1996 to 2001 among children and adolescents.[7] An insurance claims analysis that included data from all 50 states from January to December 2001 reported that antipsychotics were prescribed to 16,599 of the 6,213,824 patients 19 years of age and younger.[8] A nationwide review of the prescription patterns of office-based physicians conducted from 2000 to 2002 documented that for every 100,000 children and adolescents in the United States, 1341 were treated with antipsychotic medication.[9] Another study examining antipsychotic prescribing practices for children in the United States estimated that 2.08 million outpatient visits from 2003 to 2004, that is, 1% of all child and adolescent outpatient visits, involved the prescribing of antipsychotic medication.[10] There are indications that this growth may be most pronounced for young girls because there was a documented 110% increase in prescriptions for atypical antipsychotics written for females under the age of 20 in comparison with a 27% increase for females 20 years old and older.[6,11] However, most reports indicate that the use of antipsychotics remains far more prevalent in males.[7–10]

Results from an analysis conducted by the pharmacy benefit manager Medco Health Solutions, Inc., over a span of 4 years (2001–2005) are posted on the popular Internet site WebMD.com. These results show an 80% increase in antipsychotic prescriptions for children and adolescents versus a 46% increase among patients ages 20 to 44. According to the report, in the final year of the study, atypical antipsychotics constituted roughly 97% of the antipsychotic prescriptions for children. These children were prescribed atypical antipsychotics at a rate of 6 per 1000, a rate that was only slightly lower than that in adults (10 per 1000). Data from other sources also published on WebMD.com suggest that prescriptions for antipsychotic drugs for patients 20 years of age and younger increased from about 201,000 in 1993 to 1.2 million in 2002. A nationwide survey documented that a total of 6 million prescriptions for atypical antipsychotics were written for children from 1995 to 2002.[11]

These data clearly indicate that physicians from primary care, pediatrics, pediatric neurology, and psychiatry are currently using antipsychotic agents to treat various disorders in children much more frequently than in the near past. It is important to investigate the possible factors that may have contributed to this phenomenon. It is also important to note that the increased use of antipsychotics for children mostly reflects their use for off-label indications because there are only 2 atypical agents (eg, risperidone and aripiprazole) that have attained FDA approval for use in the treatment of particular childhood psychiatric disorders. The off-label use of any pharmacological agent is considered ethical and legal, partially because new discoveries may change the clinical practice and the standards of care faster than the FDA process of new drug approval. The FDA requires that physicians prescribing drugs for off-label treatment "must be well informed about the product, must base its use on firm scientific rationale and on sound medical evidence, and maintain records of the product's use and effects".[12] It goes without saying that any clinician who is considering prescribing antipsychotic agents for off-label use must balance evidence for the drug's clinical efficacy with information regarding its safety when making a decision. The remainder of this review presents data related to both of these issues.

FACTORS RELATED TO INCREASED USE OF ANTIPSYCHOTICS IN CHILDREN

Several factors may have contributed to the phenomenon of increased antipsychotic use in children. These include an improved understanding of childhood psychopathology and accumulating evidence supporting the notion that the use of antipsychotics in children provides clinical benefit to the patients.

Identification of Childhood Psychotic Disorders

Establishing a definitive diagnosis for childhood disorders with psychotic symptoms is challenging for several reasons. The diagnostic criteria for disorders such as schizophrenia, bipolar disorder, and major depressive disorder, as defined in the *Diagnostic and Statistical Manual of Mental Disorders*, reflect only the clinical presentations in adult populations, and little is known about the developmental trajectories of these disorders during childhood and adolescence; this renders the development of age-sensitive criteria for the disorders extremely difficult. Many clinicians consider it clinically prudent to avoid diagnosing a child with a disorder that carries a significant social stigma and a guarded prognosis. The fact that clear guidelines are not available and thus there is an increased chance for misdiagnosis serves to strengthen their prudence.

However, awareness of particular syndromes has been rapidly increasing as more precise definitions of the clinical signs of these diseases become available.[13] On the basis of existing evidence, some have advocated that schizophrenia in youth can be reliably diagnosed with the same criteria employed for adults because childhood-onset schizophrenia shares similarities with later-onset schizophrenia.[3] Results from FDA prevalence studies indicate that between 0.1% and 1% of all reported schizophrenia cases involve patients ages 13 to 17. In addition to schizophrenia, evidence suggests that there has also been a recent increase in the clinical diagnosis of bipolar disorder among young people.[14] Large private insurance databases collected from 1995 to 2000 show that the proportion of youth who received outpatient treatment for bipolar disorder increased by 67%, whereas the proportion who received inpatient treatment for bipolar disorder increased by 74%.[9] Furthermore, in a 10-year period, the prevalence of bipolar disorder diagnosis in adults increased nearly 2-fold, whereas the same prevalence in youth increased approximately 40-fold.[9] Other reports have documented that the median age of bipolar disorder onset is in the range of 19 to 23 years,[15,16] indicating that in 50% of patients, the illness starts before the age of 19. Often, children present with atypical psychotic features that do not strictly fit *Diagnostic and Statistical Manual of Mental Disorders* diagnostic criteria for a specific psychotic disorder. As a result, new diagnostic categories have been proposed, including multidimensionally impaired disorder and multiple complex developmental disorder. These disorders are commonly treated with antipsychotic agents and thus no doubt have contributed to the observed increase in the prevalence of these drugs in children and adolescents. With the increase in diagnosis, there has been an increase in academic attention devoted to pediatric bipolar disorder: the MEDLINE database retrieved 5 publications related to pediatric bipolar disorder published before 1980, 27 from 1980 to 1989, 50 from 1990 to 1999, and 227 from 2000 to 2005. Similarly, childhood bipolar disorder has become a regular feature in the popular press.[17,18] All of these developments have helped to raise clinical and public awareness and to promote treatment-seeking and implementation (including antipsychotics) for childhood psychotic disorders in general.

Efficacy of Antipsychotic Treatment for Pediatric Psychotic Disorders

To date, existing evidence supports the efficacy of risperidone, aripiprazole, and olanzapine in the treatment of childhood and adolescent schizophrenia and bipolar disorders, whereas clozapine seems to be an effective option in treatment-refractory cases.[19–21] As already mentioned, the novel agents risperidone and aripiprazole were recently approved by the FDA for the treatment of both adolescent schizophrenia and bipolar illness. The available experience on risperidone pharmacotherapy in schizophrenic children younger than 12 comes from a few case reports and open studies and suggests that it may be as efficient as or superior to older antipsychotics such as haloperidol.[13] In addition, several independent reports have documented that risperidone may be helpful in the treatment of manic symptoms in children.[22–24]

The other best studied but not FDA-approved atypical antipsychotic agent is olanzapine. A recent review of the literature reported that olanzapine has been evaluated in 4 open-label and 2 case-report studies of patients with childhood-onset schizophrenia and in 4 open-label trials and 1 double-blind trial of adolescent-onset schizophrenic patients.[25] On average, more than 50% of the patients in these studies were considered responders. One open-label trial compared both risperidone and olanzapine to haloperidol and showed that significant improvement was documented in all 3 groups after 4 weeks of treatment and at the 8-week follow-up. The largest double-blind study included 263 adult and adolescent schizophrenic patients and showed that olanzapine was superior to haloperidol when used for longer term treatment (ie, 12 weeks). Olanzapine has also been used in studies in children and adolescents diagnosed with acute mania. Results from 1 double-blind, randomized, placebo-controlled trial in subjects ages 10 to 17 showed that the mean baseline-to-endpoint change in the Young Mania Rating Scale total score was significantly greater for olanzapine-treated patients than for placebo-treated patients, with a greater proportion of olanzapine-treated patients meeting response and remission criteria (44.8% versus 18.5% and 35.2% versus 11.1%, respectively).[26] In addition, a review of 12 case reports and 1 open-label study showed that as many as 65% of the bipolar patients responded well to treatment.[25]

Little data are available for other atypical agents. Two small sample trials reported that quetiapine could be efficacious for the treatment of adolescent mania.[26,27] One double-blind, placebo-controlled, industry-sponsored trial and 1 open-label, industry-sponsored trial are underway for evaluating the efficacy of ziprasidone in adolescent schizophrenia, and 1 industry-sponsored, open-label trial on the efficacy of ziprasidone in children and adolescents with bipolar spectrum disorder has been completed, with results awaiting publication.[28] A review of 4 randomized, placebo-controlled trials showed that olanzapine, risperidone, quetiapine, and aripiprazole were all superior to placebo on the mean improvement of Young Mean Rating Scale scores in children and adolescents ages 10 to 17 (C. Correll, personal correspondence, 2008). Overall, despite the dearth of randomized, double-blind, controlled studies, an increasing body of evidence exists that supports the efficacy of these agents for the treatment of childhood and adolescent psychotic disorders.

Efficacy of Antipsychotic Treatment for Pediatric Nonpsychotic Disorders

Antipsychotic agents have been used as treatment augmentation for several nonpsychotic childhood disorders for which the FDA has approved other pharmacological agents as first-line treatment [ie, selective serotonin reuptake inhibitor agents for mood and anxiety spectrum disorders; stimulants and atomoxetine for attention-deficit hyperactivity disorder (ADHD)]. When used in this manner, antipsychotics have shown to be effective for the management of aggression, self-injurious behaviors, and severe anxiety with compromised reality testing, which in children may present as transient hallucinations, delusions, and disorganized behavior.

The ability of antipsychotic agents to diminish persistent aggression in children has been most extensively studied in individuals with developmental and/or disruptive behavior disorders (ie, autism, ADHD, oppositional defiant disorder, and conduct disorder). Three typical antipsychotics (Table 1) have received indications for the treatment of severe conduct problems in young children that may or may not be directly related to an underlying psychosis. Of these, chlorpromazine has been most widely used, especially in cases of acute agitation and aggression. With respect to atypical antipsychotics, there is a substantial body of literature that supports the efficacy of risperidone in reducing aggressive behaviors in disabled children,[29,30] and this is the only atypical agent that has been approved by the FDA for the treatment of agitation associated with autism in children ages 5 to 17.[31] Several other atypical antipsychotics (olanzapine, ziprasidone, quetiapine, and aripiprazole) have shown some efficacy in improving aggressiveness, hyperactivity, and self-injurious behavior in autistic children. Most data have been collected on the use of olanzapine, followed by quetiapine and ziprasidone, whereas a large network study, evaluating aripiprazole in combination with behavioral interventions in children with autism, is in progress.[32]

Aggressive behaviors are also frequently present in the clinical picture of disruptive behavior disorders, and several studies have directly investigated the effects of FDA-approved treatments (ie, stimulants) in impulsive aggression within the context of ADHD.[33] However, there is evidence that the off-label use of antipsychotics in children with ADHD has increased, as suggested by a report[6] that 43% of the prescriptions for antipsychotics for children on Medicaid in the state of Tennessee during a 5-year period (1996 to 2001) were for children with ADHD as the primary diagnosis. In comparison, children diagnosed with schizophrenia and bipolar disorder received 9% and 14%, respectively, of the antipsychotic prescriptions for the same time period. It should be emphasized that there is a significant clinical overlap between childhood bipolar illness and the childhood disruptive behavior disorders because they all share high rates of irritability, chronicity, aggressive behaviors, and mood liability.[26] The efficacy and safety of stimulants in the treatment of children with comorbid ADHD/bipolar disorder has been suggested by some[26]; however, ADHD patients with partial response to stimulants may require treatment augmentation with other agents, including antipsychotics. Of these, risperidone has been the most studied, and its efficacy for the treatment of aggressive behaviors has been documented in children with ADHD and comorbid treatment-resistant aggression,[33] disruptive behavior disorder/conduct disorder,[34–36] or mental retardation.[37,38] There are also a few case reports documenting the usefulness of olanzapine in the management of conduct problems in ADHD children.[25]

Self-injurious behaviors in youth could be a part of the clinical presentation of adolescent trichotillomania, mixed bipolar illness, or borderline personality traits and may also be the basis for the initiation of an antipsychotic medication trial.[39,40] Self-injurious behaviors may become particularly challenging to manage in developmentally and cognitively delayed children and adolescents, who may inflict injuries to themselves and others in the context of low frustration and severe agitation due to relatively minor stressors. As mentioned previously, the use of antipsychotics in such cases has shown clinical benefits.

Other types of childhood disorders for which there are published data (primarily case reports) suggesting the usefulness of antipsychotic agents include childhood obsessive-compulsive disorder,[41] posttraumatic stress disorder,[42,43] and anorexia nervosa.[25] The presence of psychotic symptoms (ie, obsessions, delusions, agitation, and anxiety) has been cited as the main therapeutic reason for using antipsychotics to treat these disorders, for these symptoms become less intense after the initiation of the antipsychotics. In the case of anorexia, the ability of olanzapine to stimulate patients' appetite (an otherwise problematic side effect discussed in more detail in the next section) has resulted in successful weight restoration.

ADVERSE EFFECTS ASSOCIATED WITH ANTIPSYCHOTIC USE IN CHILDREN

Some of the most frequently reported side effects associated with the treatment of antipsychotic agents are summarized in Table 2.

Neuromotor Side Effects

As discussed previously, the high affinity of the typical antipsychotics to the D2 dopamine receptor may present as some worrisome adverse reactions. Extrapyramidal side effects such as Parkinsonism, acute dystonia, akathisia, tardive dyskinesia, and withdrawal dyskinesia have been well recognized in adults. Because of their novel neurotransmitter profiles (ie, serotonergic agonistic effects for most new agents and partial dopaminergic agonistic effects for aripiprazole), the atypical antipsychotics are considered less likely to produce neuromotor reactions. It should be pointed out that children are more likely to develop these types of problems than adults. Several reports have documented high levels of extrapyramidal side effects with haloperidol, risperidone, and olanzapine, whereas clozapine and quetiapine have been associated with relatively low extrapyramidal side effect rates in children.[4] One report suggests that, when slowly titrated, the partial agonist agent aripiprazole may reduce rates of akathisia,[44] and data are still needed for ziprasidone. Pediatric patients are also at risk of developing withdrawal dyskinesias associated with a discontinuation of an antipsychotic, but unlike in adults, these are mostly reversible. Last but not least, a meta-analysis of 10 pediatric studies (a total of 783 patients) reported only 3 cases of tardive dyskinesia in children; this is a rate approximately half of the rate documented by another meta-analysis in 1964 adults.[4] At present, the incidence rate of these problems associated with long-term treatment in children remains unknown.

Table 2. *Side Effects Associated with the Chronic Use of Antipsychotics and Suggested Monitoring Procedures.*

Type of Adverse Effect	Symptomathology	Monitoring at Baseline	Monitoring at Follow-Up
Neuromotor/extrapyramidal	Parkinsonism	SARS and ESRS	SARS and ESRS during titration, at 3 months, and then annually
	Akathisia	SARS and ESRS	Same as above
	Tardive dyskinesia	AIMS	Repeat at 3 months and annually.
Metabolic syndrome	Weight gain	Measurement of height and weight and calculation of BMI percentile (Z score)	Measurement of height and weight and calculation of BMI percentile at each visit
	Dyslipidemia	Lipid profile (fasting triglycerides, LDL, and HDL)	Repeat lipid profile at 3 months and then every 6 months.
	Hypertension	Vital signs (heart rate and blood pressure)	Repeat at 3 months and then annually.
Diabetes mellitus	Polyuria Polydipsia Unintended weight loss	Fasting glucose and fasting triglyceride/HDL ratio	Repeat fasting glucose and fasting triglyceride/HDL ratio at 3 months and then every 6 months.
Prolactin-related	Hypogonadism Galactorrhea Decreased libido	Prolactin level	Repeat if symptoms develop.
	Osteoporosis		Bone density scan may be indicated
Cardiac side effects	Dizziness/orthostasis	Vital signs (heart rate and blood pressure)	Repeat at 3 months and then annually.
	QTc prolongation	ECG if the patient is taking ziprasidone or clozapine	ECG during titration and at the maximum dose if the patient is taking ziprasidone
	Myocarditis		ECG looking for T wave inversions if symptoms develop
Liver toxicity	Abdominal pain and jaundice	Liver function test	Repeat at 3 months and then annually.

Abbreviations: AIMS, Abnormal Involuntary Movement Scale; BMI, body mass index; ECG, electrocardiogram; ESRS, Extrapyramidal Symptom Rating Scale; HDL, high-density lipoprotein; LDL, low-density lipoprotein; SARS, Simpson Angus Rating Scale.

Prolactin-Induced Side Effects

Because dopamine regulates the prolactin secretion from the pituitary lactotrophic cells, the D2-receptor antagonistic effect of antipsychotic medications may result in blocking of the tonic inhibitory effect of dopamine, leading to increased levels of serum prolactin. The relative potency of antipsychotic drugs in inducing hyperprolactinemia is as follows: risperidone > haloperidol > olanzapine > ziprasidone > quetiapine > clozapine > aripiprazole.[45] Both adults and children may develop hyperprolactinemia in response to antipsychotics, with some evidence suggesting that this effect may be most pronounced in postpubertal youth.[46] Hyperprolactinemia may result in hypogonadism (amenorrhea, low estrogen in females, and low testosterone in males), galactorrhea, decreased libido, erectile dysfunction, osteoporosis, and possible delay in pubertal maturation. However, some data indicate that serum prolactin levels are not tightly correlated with these side effects and that not all patients with hyperprolactinemia will develop these types of symptoms.[47] Not last, a recent report has raised concern that the propensity of a medication to cause hyperprolactinemia may be related to a risk for pituitary tumors.[48] As this potentially worrisome association needs to be clarified further, it should be mentioned that silent pituitary tumors are found in 3% to 27% of autopsies and pituitary magnetic resonance imaging scans, and this suggests that these benign tumors may have a high incidence rate independent of the antipsychotic treatment.[49,50]

Weight Gain

Although all of the atypical antipsychotics share a tendency to promote weight gain, especially in drug-naïve patients, larger, long-term studies are needed to better define the relative risks of weight gain in children and adolescents. Existing data from adults[51] and from an ongoing, large-scale naturalistic study in children and adolescents[52] propose the following rank order in terms of ability to promote weight gain and development of the metabolic syndrome: clozapine = olanzapine > risperidone ≥ quetiapine > ziprasidone ≥ aripiprazole.

Weight gain is not an isolated problem; it has been associated with lipid abnormalities, such as elevated triglyceride, total cholesterol, and low-density lipoprotein–cholesterol levels and/or decreased high-density lipoprotein (HDL)–cholesterol levels, in adults[53] and in youths.[54] Steatohepatitis and pancreatitis are other potential complications of weight gain. Moreover, weight gain and obesity are also associated with metabolic syndrome, a constellation of physical and laboratory features that is more common in obese patients. When antipsychotics were coadministered with stimulants, the weight loss–inducing effect of the latter did not balance out the weight gain induced by former.[55] Furthermore, cotreatment with atypical antipsychotics and mood stabilizers had a cumulative effect, so that the weight gain observed was greater than the weight gain typically observed in patients treated with only one of the medications.[4]

Metabolic Syndrome

The clinical features of metabolic syndrome include abdominal obesity, dyslipidemia (discussed previously), glucose intolerance, and hypertension. A common cause for all features of metabolic syndrome appears to be insulin resistance, which in turn is also associated with weight gain. Metabolic syndrome occurs in about 5% to 10% of all adolescents in the US population and in greater than 30% of those who are overweight.[56,57] Importantly, the occurrence of metabolic syndrome in young individuals predicts early atherosclerosis and vascular disease in adulthood.[58,59] As metabolic syndrome has been found to be more prevalent in adults treated with atypical antipsychotics in comparison with healthy controls[60,61] studies on the prevalence of metabolic syndrome in children and adolescents treated with antipsychotics must be pursued. The relationship between atypical antipsychotics and metabolic syndrome may be difficult to establish because other factors, such as the naturalistic course of the illness, genetic vulnerability, and an unhealthy lifestyle, may also play a role.[62,63] Nevertheless, because weight gain is a leading pathway to metabolic syndrome, this risk needs to be considered in youngsters receiving antipsychotic medications, and these patients should be monitored accordingly.

Diabetes Mellitus

Diabetes mellitus is another result of significant weight gain, and the FDA has issued a black-box warning about the development of diabetes mellitus in patients receiving any atypical antipsychotic. The development of diabetes associated with antipsychotic treatments has been described in children and adolescents. Patients at elevated risk include those with a family history of diabetes, nonwhite ethnic groups, patients who rapidly gain weight on the antipsychotic drug, and patients receiving either olanzapine or clozapine.[64] In addition to the known contribution of weight gain to the development of diabetes, it has been hypothesized that atypical antipsychotics may also cause the disorder by directly affecting insulin secretion.[65,66]

Other Side Effects

One of the most commonly reported side effects of antipsychotic treatment is sedation, with rates ranging from 0% to 33% for aripiprazole, from 42% to 69% for ziprasidone, from 25% to 80% for quetiapine, from 29% to 89% for risperidone, from 44% to 94% for olanzapine, and from 46% to 90% for clozapine.[4] Orthostasis and dizziness, related to alpha-1 receptor blockade, seem to be most pronounced with clozapine and quetiapine. The antipsychotic agents were also found to prolong the heart rate–corrected QTc interval on electrocardiogram, and this can potentially lead to fatal arrhythmias.[67] This phenomenon has been documented in a small number of children, and its clinical significance is unclear. Risk for more severe reactions such as myocarditis, neutropenia, and agranulocytosis has been associated exclusively with clozapine treatment.

MONITORING STRATEGIES

Several publications have reviewed the issue of monitoring of patients on atypical antipsychotic agents,[45] and some of the recommended algorithms are summarized on Table 2. A baseline assessment should include measurements of body height and weight, a body mass index should be calculated, and these can be repeated at each visit. Although a baseline electrocardiogram is recommended only for particular agents (see Table 2), vital signs including heart rate and blood pressure should be routinely assessed with all antipsychotic treatments. Quarterly age-adjusted and sex-adjusted waist circumference percentiles should be measured in children and adolescents who are still growing.[68] If possible, a baseline fasting blood glucose measurement should be obtained before an antipsychotic drug is prescribed; follow-up blood glucose determinations should be performed 3 months after the drug is started and every 6 months thereafter. High-risk patients (as discussed previously) should have fasting blood glucose measurements performed at least quarterly. Clinicians should diligently inquire about unintended weight loss, polyuria, and polydipsia, all of which could indicate the onset of hyperglycemia. In conjunction with fasting blood sugar measurements, a fasting serum lipid panel should be obtained at baseline, at 3 months, and every 6 to 12 months thereafter if results are within normal limits and body mass index percentile values remain stable. A compensatory increase of insulin secretion in response to increasing body fat mass may indicate insulin resistance, which is considered an earlier indicator of risk for diabetes.[69] The direct calculation for insulin resistance is expensive and impractical. Alternatively, a fasting triglycerides to HDL-cholesterol ratio of 3.5 mg/dL or higher has been proposed as a widely applicable and sensitive measure of insulin resistance.[70] However, this use of the triglyceride/HDL-cholesterol ratio has yet to be replicated in other adult populations and to be validated in children and adolescents.

CONCLUSION

Improved diagnostic and treatment algorithms for childhood psychotic disorders have contributed to a wider use of antipsychotic medications, particularly of the novel atypical agents. Their efficacy as first-line treatments for childhood psychotic disorders and therapeutic augmentation of nonpsychotic conditions has been supported by evidence from an increasing number of clinical trials. So far, 2 of the atypical agents (ie, risperidone and aripiprazole) have also received FDA approval for the treatment of schizophrenia, bipolar disorder, and agitation associated with autism in youth. However, the clinical benefits of such treatments appear to come at the expense of acute and chronic adverse effects, some of which may cause long-term consequences, especially when used early in life. Side effects such as prolactin-induced hypogonadism and decreased libido, weight gain, dyslipidemia, and diabetes mellitus have raised concerns about the increased use of these agents in children. Until more definitive evidence becomes available, the clinician's decision for the initiation of antipsychotic treatment for pediatric patients should be carefully justified and based on a thorough psychiatric assessment, a consideration of alternative treatments, and diligent monitoring of the treatment process with respect to both clinical efficacy and safety.

ACKNOWLEDGMENT

This work was supported by NIDA/AACAP K23 (PA-00–003).

DISCLOSURES

There is nothing to report.

REFERENCES

1. Remschmidt H, Theisen FM. Schizophrenia and related disorders in children and adolescents. *J Neural Transm Suppl* 2005; 69: 121–141.
2. Zalsman G, Oquendo MA, Greenhill L, et al. Neurobiology of depression in children and adolescents. *Child Adolesc Psychiatr Clin N Am* 2006; 15: 843–868, vii–viii.
3. Asarnow JR, Tompson MC, McGrath EP. Annotation: childhood-onset schizophrenia: clinical and treatment issues. *J Child Psychol Psychiatry* 2004; 45: 180–194.
4. Correll C. Antipsychotic use in children and adolescents: minimizing adverse effects to maximize outcomes. *J Am Acad Child Adolesc Psychiatry* 2008; 47: 9–20.
5. Kane JM. Extrapyramidal side effects are unacceptable. *Eur Neuropsychopharmacol* 2001; 11(suppl 4): S397–S403.
6. Cooper WO, Hickson GB, Fuchs C, et al. New users of antipsychotic medications among children enrolled in TennCare. *Arch Pediatr Adolesc Med* 2004; 158: 753–759.
7. Patel NC, Crismon ML, Hoagwood K, et al. Trends in the use of typical and atypical antipsychotics in children and adolescents. *J Am Acad Child Adolesc Psychiatry* 2005; 44: 548–556.
8. Curtis LH, Masselink LE, Ostbye T, et al. Prevalence of atypical antipsychotic drug use among commercially insured youths in the United States. *Arch Pediatr Adolesc Med* 2005; 159: 362–366.
9. Olfson M, Blanco C, Liu L, et al. National trends in the outpatient treatment of children and adolescents with antipsychotic drugs. *Arch Gen Psychiatry* 2006; 63: 679–685.
10. Aparasu RR, Bhatara V, Gupta S. U.S. national trends in the use of antipsychotics during office visits, 1998–2002. *Ann Clin Psychiatry* 2005; 17: 147–152.

11. Cooper WO, Arbogast PG, Ding H, et al. Trends in prescribing of antipsychotic medications for US children. *Ambul Pediatr* 2006; 6: 79–83.
12. Food and Drug Administration. Available at: http://www.fda.gov. Accessed May 2008.
13. Masi G, Perugi G, Millepiedi S, et al. Clinical implications of DSM-IV subtyping of bipolar disorders in referred children and adolescents. *J Am Acad Child Adolesc Psychiatry* 2007; 46: 1299–1306.
14. Leverich GS, Post RM, Keck PE Jr, et al. The poor prognosis of childhood-onset bipolar disorder. *J Pediatr* 2007; 150: 485–490.
15. Grant BF, Stinson FS, Hasin DS, et al. Prevalence, correlates, and comorbidity of bipolar I disorder and axis I and II disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *J Clin Psychiatry* 2005; 66: 1205–1215.
16. Kessler RC, Angermeyer M, Anthony JC, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry* 2007; 6: 168–176.
17. Kluger J. Young and bipolar. Available at: http://www.time.com/time/magazine/article/0,9171,1003048-1,00.html. Accessed May 2008.
18. CBS Evening News, Misdiagnosing bipolar kids, November 11, 2003.
19. Kumra S, Kranzler H, Gerbino-Rosen G, et al. Clozapine and "high-dose" olanzapine in refractory early-onset schizophrenia: a 12-week randomized and double-blind comparison. *Biol Psychiatry* 2008; 63: 524–529.
20. Shaw P, Sporn A, Gogtay N, et al. Childhood-onset schizophrenia: a double-blind, randomized clozapine-olanzapine comparison. *Arch Gen Psychiatry* 2006; 63: 721–730.
21. Gogtay N, Rapoport J. Clozapine use in children and adolescents. *Expert Opin Pharmacother* 2008; 9: 459–465.
22. Frazier JA, Meyer MC, Biederman J, et al. Risperidone treatment for juvenile bipolar disorder: a retrospective chart review. *J Am Acad Child Adolesc Psychiatry* 1999; 38: 960–965.
23. Simeon J, Milin R, Walker S. A retrospective chart review of risperidone use in treatment-resistant children and adolescents with psychiatric disorders. *Prog Neuropsychopharmacol Biol Psychiatry* 2002; 26: 267–275.
24. Biederman J, Mick E, Faraone SV, et al. Risperidone for the treatment of affective symptoms in children with disruptive behavior disorder: a post hoc analysis of data from a 6-week, multicenter, randomized, double-blind, parallel-arm study. *Clin Ther* 2006; 28: 794–800.
25. Fremaux T, Reymann JM, Chevreuil C, Bentue-Ferrer D. Prescription of olanzapine in children and adolescent psychiatric patients. *Encephale* 2007; 33: 188–196.
26. Tohen M, Kryzhanovskaya L, Carlson G, et al. Olanzapine versus placebo in the treatment of adolescents with bipolar mania. *Am J Psychiatry* 2007; 164: 1547–1556.
27. Delbello MP, Schwiers ML, Rosenberg HL, Strakowski SM. A double-blind, randomized, placebo-controlled study of quetiapine as adjunctive treatment for adolescent mania. *J Am Acad Child Adolesc Psychiatry* 2002; 41: 1216–1223.
28. Greenberg WM, Citrome L. Ziprasidone for schizophrenia and bipolar disorder: a review of the clinical trials. *CNS Drug Rev* 2007; 13: 137–177.
29. Myers SM. The status of pharmacotherapy for autism spectrum disorders. *Expert Opin Pharmacother* 2007; 8: 1579–1603.
30. Stachnik JM, Nunn-Thompson C. Use of atypical antipsychotics in the treatment of autistic disorder. *Ann Pharmacother* 2007; 41: 626–634.
31. Scahill L, Koenig K, Carroll DH, Pachler M. Risperidone approved for the treatment of serious behavioral problems in children with autism. *J Child Adolesc Psychiatr Nurs* 2007; 20: 188–190.
32. King BH, Bostic JQ. An update on pharmacologic treatments for autism spectrum disorders. *Child Adolesc Psychiatr Clin N Am* 2006; 15: 161–175.
33. Armenteros JL, Lewis JE, Davalos M. Risperidone augmentation for treatment-resistant aggression in attention-deficit/hyperactivity disorder: a placebo-controlled pilot study. *J Am Acad Child Adolesc Psychiatry* 2007; 46: 558–565.
34. Gunther T, Herpertz-Dahlmann B, Jolles J, Konrad K. The influence of risperidone on attentional functions in children and adolescents with attention-deficit/hyperactivity disorder and co-morbid disruptive behavior disorder. *J Child Adolesc Psychopharmacol* 2006; 16: 725–735.
35. Turgay A. Treatment of comorbidity in conduct disorder with attention-deficit hyperactivity disorder (ADHD). *Essent Psychopharmacol* 2005; 6: 277–290.
36. Tcheremissine OV, Lieving LM. Pharmacological aspects of the treatment of conduct disorder in children and adolescents. *CNS Drugs* 2006; 20: 549–565.
37. Correia Filho AG, Bodanese R, Silva TL, et al. Comparison of risperidone and methylphenidate for reducing ADHD symptoms in children and adolescents with moderate mental retardation. *J Am Acad Child Adolesc Psychiatry* 2005; 44: 748–755.
38. Aman M, Buitelaar J, Smedt GD, et al. Pharmacotherapy of disruptive behavior and item changes on a standardized rating scale: pooled analysis of risperidone effects in children with subaverage IQ. *J Child Adolesc Psychopharmacol* 2005; 15: 220–232.
39. Segal-Trivitz Y, Bloch Y, Goldburt Y, et al. Comparison of symptoms and treatments of adults and adolescents with borderline personality disorder. *Int J Adolesc Med Health* 2006; 18: 215–220.
40. Pathak S, Danielyan A, Kowatch RA. Successful treatment of trichotillomania with olanzapine augmentation in an adolescent. *J Child Adolesc Psychopharmacol* 2004; 14: 153–154.
41. Grados MA, Riddle MA, Samuels JF, et al. The familial phenotype of obsessive-compulsive disorder in relation to tic disorders: the Hopkins OCD family study. *Biol Psychiatry* 2001; 50: 559–565.
42. Hamner MB, Robert S. Emerging roles for atypical antipsychotics in chronic post-traumatic stress disorder. *Expert Rev Neurother* 2005; 5: 267–275.
43. Donnelly CL. Pharmacologic treatment approaches for children and adolescents with posttraumatic stress disorder. *Child Adolesc Psychiatr Clin N Am* 2003; 12: 251–269.
44. Robb AJ, Findling RL, Nyilas M. et al. Efficacy of aripiprazole in the treatment of adolescents with schizophrenia. Poster presented at: 160th Annual

Meeting of the American Psychiatric Association; May 19–24, 2007; San Diego, CA.

45. Correll CU, Carlson HE. Endocrine and metabolic adverse effects of psychotropic medications in children and adolescents. *J Am Acad Child Adolesc Psychiatry* 2006; 45: 771–791.
46. Wudarsky M, Nicolson R, Hamburger SD, et al. Elevated prolactin in pediatric patients on typical and atypical antipsychotics. *J Child Adolesc Psychopharmacol* 1999; 9: 239–245.
47. Masi G, Cosenza A, Mucci M, Brovedani P. A 3-year naturalistic study of 53 preschool children with pervasive developmental disorders treated with risperidone. *J Clin Psychiatry* 2003; 64: 1039–1047.
48. Szarfman A, Tonning JM, Levine JG, Doraiswamy PM. Atypical antipsychotics and pituitary tumors: a pharmacovigilance study. *Pharmacotherapy* 2006; 26: 748–758.
49. Chanson P, Young J. Pituitary incidentalomas. *Endocrinologist* 2003; 13: 124–135.
50. Oyama K, Sanno N, Tahara S, Teramoto A. Management of pituitary incidentalomas: according to a survey of pituitary incidentalomas in Japan. *Semin Ultrasound CT MR* 2005; 26: 47–50.
51. Casey DE, Haupt DW, Newcomer JW, et al. Antipsychotic-induced weight gain and metabolic abnormalities: implications for increased mortality in patients with schizophrenia. *J Clin Psychiatry* 2004; 65(suppl 7): 4–18; quiz 19–20.
52. Correll CU, Parikh UH, Mughal T, et al. Body composition changes associated with second-generation antipsychotics. *Biol Psychiatry* 2005; 36: 57.
53. Lindenmayer JP, Czobor P, Volavka J, et al. Changes in glucose and cholesterol levels in patients with schizophrenia treated with typical or atypical antipsychotics. *Am J Psychiatry* 2003; 160: 290–296.
54. Correll CU, Parikh UH, Mughal T, et al. New-onset dyslipidemia in antipsychotic-naïve youngsters treated with atypical antipsychotics. *Biol Psychiatry* 2005; 36: 7.
55. Aman MG, Binder C, Turgay A. Risperidone effects in the presence/absence of psychostimulant medicine in children with ADHD, other disruptive behavior disorders, and subaverage IQ. *J Child Adolesc Psychopharmacol* 2004; 14: 243–254.
56. de Ferranti SD, Gauvreau K, Ludwig DS, et al. Prevalence of the metabolic syndrome in American adolescents: findings from the Third National Health and Nutrition Examination Survey. *Circulation* 2004; 110: 2494–2497.
57. Duncan GE, Li SM, Zhou XH. Prevalence and trends of a metabolic syndrome phenotype among U.S. adolescents, 1999–2000. *Diabetes Care* 2004; 27: 2438–2443.
58. Li S, Chen W, Srinivasan SR, et al. Childhood cardiovascular risk factors and carotid vascular changes in adulthood: the Bogalusa Heart Study. *JAMA* 2003; 290: 2271–2276.
59. Raitakari OT, Taimela S, Porkka KV, Viikari JS. Effect of leisure-time physical activity change on high-density lipoprotein cholesterol in adolescents and young adults. *Ann Med* 1996; 28: 259–263.
60. McEvoy JP, Meyer JM, Goff DC, et al. Prevalence of the metabolic syndrome in patients with schizophrenia: baseline results from the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) schizophrenia trial and comparison with national estimates from NHANES III. *Schizophr Res* 2005; 80: 19–32.
61. Straker D, Correll CU, Kramer-Ginsberg E, et al. Cost-effective screening for the metabolic syndrome in patients treated with second-generation antipsychotic medications. *Am J Psychiatry* 2005; 162: 1217–1221.
62. Mackin P, Watkinson HM, Young AH. Prevalence of obesity, glucose homeostasis disorders and metabolic syndrome in psychiatric patients taking typical or atypical antipsychotic drugs: a cross-sectional study. *Diabetologia* 2005; 48: 215–221.
63. Toalson P, Ahmed S, Hardy T, Kabinoff G. The metabolic syndrome in patients with severe mental illnesses. *Prim Care Companion J Clin Psychiatry* 2004; 6: 152–158.
64. Henderson DC. Clinical experience with insulin resistance, diabetic ketoacidosis, and type 2 diabetes mellitus in patients treated with atypical antipsychotic agents. *J Clin Psychiatry* 2001; 62(suppl 27): 10–14; discussion 40–41.
65. Ader M, Kim SP, Catalano KJ, et al. Metabolic dysregulation with atypical antipsychotics occurs in the absence of underlying disease: a placebo-controlled study of olanzapine and risperidone in dogs. *Diabetes* 2005; 54: 862–871.
66. Henderson DC, Cagliero E, Copeland PM, et al. Glucose metabolism in patients with schizophrenia treated with atypical antipsychotic agents: a frequently sampled intravenous glucose tolerance test and minimal model analysis. *Arch Gen Psychiatry* 2005; 62: 19–28.
67. Blair J, Scahill L, State M, Martin A. Electrocardiographic changes in children and adolescents treated with ziprasidone: a prospective study. *J Am Acad Child Adolesc Psychiatry* 2005; 44: 73–79.
68. Fernandez JR, Redden DT, Pietrobelli A, Allison DB. Waist circumference percentiles in nationally representative samples of African-American, European-American, and Mexican-American children and adolescents. *J Pediatr* 2004; 145: 439–444.
69. Rosenbloom AL, Joe JR, Young RS, Winter WE. Emerging epidemic of type 2 diabetes in youth. *Diabetes Care* 1999; 22: 345–354.
70. McLaughlin T, Reaven G, Abbasi F, et al. Is there a simple way to identify insulin-resistant individuals at increased risk of cardiovascular disease? *Am J Cardiol* 2005; 96: 399–404.

CONTROVERSY

Predictive Genetic Testing in Minors for Adult-Onset Genetic Diseases

Pascal Borry, Tom Goffin, Herman Nys and Kris Dierickx

Centre for Biomedical Ethics and Law, Katholieke Universiteit Leuven, Leuven, Belgium

ABSTRACT

This article analyses the ethical discussion of predictive genetic testing in minors for genetic conditions for which there is no phenotypic evidence for disease at the considered time of testing and for which there is currently no treatment available to prevent or forestall the development of the condition. After a presentation of the position of various professional guidelines, we discuss the position and arguments that have been advanced in a recent article that defends a position that is opposed to the professional recommendations. In the article, we discuss the position of voluntary choices and autonomous and informed decision-making in a context of open communication. Thereafter, we analyze the nonmedical benefits and harm related to this type of testing. Finally, we critically analyze 4 arguments: the potential provision of good news if a test is performed, the unbearability of knowing, identity and adjustment, and parental anxiety and uncertainty. ***Mt Sinai J Med 75:287–296, 2008.*** *© 2008 Mount Sinai School of Medicine*

Key Words: adult-onset genetic diseases, genetic diseases, genetic testing, predictive testing.

Presymptomatic and predictive genetic testing makes it possible to trace a genetic defect before the manifestation of symptoms. On the one hand, this offers the advantage that individuals who carry a genetic risk can find out more concerning their future health risks, whereas they remained uncertain in the past. On the other hand, this has led to ethical discussions about whether or not a certain test should be carried out and also about how and in what manner this test should be carried out. Predictive genetic tests may have far-reaching consequences for test applicants, their family members, and society.[1] As a result, ethical reflection[2,3] has always been concerned with the careful transfer of information about genetic tests and test results, the confidentiality of genetic information, the voluntariness of requesting a genetic test, the responsibility toward blood relatives, and the (potential) psychosocial impact of a test on the applicant.[4] An even more cautious approach has been envisaged when such testing is considered in children and adolescents. This originates from the fear that testing in childhood or adolescence could create devastating social, emotional, psychosocial, and educational consequences in the child or adolescent.

In this article, we analyze the ethical discussion of predictive genetic testing for genetic conditions for which there is no phenotypic evidence for disease at the time of testing and for which there is currently no treatment available to prevent or forestall the development of the condition. In particular, we discuss the policies for predictive testing of minors for adult-onset genetic conditions (eg, Huntington's disease and polycystic kidney disease). To avoid confusion in terminology, by *minors* we mean all persons who have not reached the age of majority (which is in most countries 18 years).

In the first part of the article, we present the positions of various genetic and medical guidelines with respect to predictive genetic testing of minors for adult-onset diseases. In the second part of the article, we discuss how the guidelines have emphasized that requests for predictive genetic tests for adult-onset diseases should be voluntary and should come from

Address Correspondence to:

Pascal Borry
Centre for Biomedical Ethics
and Law
Katholieke Universiteit Leuven
Kapucijnenvoer 35/3, Box 700
3000 Leuven, Belgium
Email: pascal.borry@
med.kuleuven.be

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20038

© 2008 Mount Sinai School of Medicine

a competent applicant, without any form of pressure. In the third part of this article, we analyze how the perspective of nonmedical benefits and harm is affecting the discussion on this type of testing. In the article, we refer to and comment on the positions and arguments that have been used in an article of Rhodes[5] that was recently published in this journal. In that article, the author defends a position that is "diametrically opposed to the policies and guidelines of the genetics and pediatric communities." Rhodes concluded that

> Pediatricians and geneticists should encourage parents to pursue genetic testing of children at a young age. . . . Pediatricians and geneticists should encourage reluctant parents to have their children tested–and then accept the parents' decision.

In light of this article by Rhodes, a more extensive and elaborate debate on the issue of predictive genetic testing in minors for adult-onset genetic diseases seemed appropriate to us.

POLICIES, GUIDELINES, AND RECOMMENDATIONS

In the last decade, various genetic and medical associations have published guidelines[6–13] regarding predictive genetic testing in minors. All guidelines clearly put forward that medical benefit is the main justification for testing minors for adult-onset diseases. They stress in a similar way that predictive genetic testing should be performed only for persons of legal age, except for disorders for which preventive or therapeutic actions could be initiated. They clearly advance in similar words that, when they are talking specifically about predictive and presymptomatic tests for late-onset disorders, such testing is recommended only when "established, effective, and important medical treatment"[14] can be offered or when testing "provides scope for treatment which to any essential degree prevents, defers or alleviates the outbreak of disease or the consequences of the outbreak of disease."[15] The rationale is that medical benefit to the child should be the primary justification for genetic testing in children and adolescents. The availability of preventive or other therapeutic measures leads in some cases even to the requirement to perform a genetic test.[13] Also, some disease-specific guidelines have been elaborated and put forward a similar position. In the case of Huntington's disease, for example, the International Huntington Association and the World Federation of Neurology already published in 1990 a policy statement concerning the provision of the test to minors. In the revised version of 1994, the text still kept the provision that the test should be available only to "those having reached the age of majority."

Although in most countries a person's 18th birthday draws the line between childhood and adulthood, most clinical-ethical guidelines do not use this demarcation as a strict rule for accepting or refusing genetic testing for adult-onset diseases. In the case of late-onset diseases, it is emphasized that age "should be given flexible consideration"[16] and that genetic tests should be deferred as much as possible to the moment adolescents are "competent to make a free informed decision."[17] The rationale behind this option is that predictive testing for adult-onset diseases "should be delayed until the person is old enough to make an informed choice."[18] "The degree of maturity and state of development"[14] and "conditions of competence, voluntariness, and adequate understanding of information"[6] are considered crucial criteria in the assessment of adolescents' requests for genetic testing. Clinical-ethical guidelines should of course always be assessed in the particular legal context in which they have been elaborated. Legal provisions on the age of medical majority[19] contextualize the professional recommendations on the age at which minors can consent to predictive genetic testing; for example, it is clear that Danish law on patients' rights[20] supports the professional recommendation that competent patients between 15 and 18 years may personally give informed consent to genetic testing after the person holding parental authority has been informed and allowed to participate in the decision-making. These clinical-ethical guidelines stress that children should receive education and counseling appropriate to their age and maturity. Postponing testing is not considered a reason to avoid discussing the issues of genetic risk and/or genetic testing with younger children.[10]

PERSONAL REQUEST

Predictive genetic tests provide information about the medical future of a healthy person. In this sense, there is a substantial difference between a predictive genetic test and a traditional medical diagnosis that says something about the current medical condition of the patient. The person who undergoes a predictive test is not (yet) a patient. In other words, it concerns a real presymptomatic diagnosis. In general, there is no urgent medical reason to carry out a predictive genetic test. Because such a test may have a dramatic impact, it has

DOI:10.1002/MSJ

been stressed that it is of utmost importance for the request for a test to be voluntary and to come from a competent applicant, without any form of external pressure (eg, from a partner or family).[21] In adults, a multidisciplinary team (with clinical geneticists, psychologists, psychiatrists, neurologists, and social workers) evaluates applications for a predictive genetic test for a condition. During pretest counseling, adequate information is provided about the disorder, the predictive test, and the possible psychological impact of a test on the person.[21] People who take part in a predictive genetic test should be clearly informed about not only the possibilities but also the limitations and substantial elements of uncertainty that are linked to the test. Such a multidisciplinary approach is necessary to safeguard the well-being of the test applicant and to protect him against the possible pernicious impact of insufficiently prepared tests.

In this light, professional guidelines have always defended the position of voluntary choices and autonomous and informed decision-making in a context of open discussion about familial hereditary diseases. Also, in the case of predictive genetic testing in minors, the option has been made to let the future mature patient decide about such a test instead of letting parents decide without children's consent or assent. In a recent article, Rhodes[5] challenges this view and questions why this is the case whereas "autonomy is not an issue in medical decisions made on behalf of children whenever doctors and parents impose vaccinations, antibiotic injections or other treatment on children who may scream their opposition." She refers to the fact that genetic diseases are familial diseases and that some family members are already afflicted by the disease. The environment in which the child is growing is "an unavoidable piece of the child's inheritance." As a result, she argues that the decision whether to raise the child with knowledge or ignorance of her or his genetic status is one that the family must make. She argues as well that it does not violate the principle of confidentiality if a minor is being tested on request of the parents. On the basis of parental authority, she argues that parents should be authorized to decide whether to take a sample of their child's blood, have the genetic test performed, and be provided with the genetic information of their children:

> Parental authority is the model for making medical decisions on behalf of children. This is a well-accepted social norm that is supported by our laws, and it is usually the most appropriate mechanism for making decisions on behalf of children. Parents are typically deeply concerned with their children's well-being.

However, Rhodes[5] fails in her position to give a place to minors and to make a distinction between predictive genetic testing of minors and any kind of other medical procedure. We discuss both observations.

First, because young children are not able to make complex decisions for themselves, the authority to make medical decisions on behalf of a child usually falls to the child's parents. Parents (or legal guardians) have indeed received ethical and legal authority over their children's medical care. This includes the right to initiate, refuse, or stop preventive or therapeutic measures. The basis of this position is that parents are in a unique moral relationship with their children and are expected to make decisions that are in the best interests of their children, that is, to act "so as to promote maximally the good of the individual."[22] Parents' ethical and legal authority over their child's medical care should be directed toward the promotion of their children's health-related interests. International declarations,[23,24] various policy documents, and publications[25] have emphasized that the greater a child's cognitive and social abilities are, the more seriously the preferences of the minor should be taken into consideration. As soon as children are able to communicate and participate in decisions that affect them, they should be encouraged to participate in all aspects of the decision-making process. They should be properly informed about medical issues that affect them, and they should be able to express their views, ask questions, or communicate their worries. Rutter[26] expressed this as follows:

> The implication is that health care professionals should be expected to talk with children, even during the pre-school period, about medical matters that involve them. This needs to be done in ways that the children can understand but that are no different in principle from the exactly comparable need to do so with respect to adults. Thus, most adults can scarcely be expected to understand all the complex considerations that go into decisions about different forms of medical treatment, or different surgical procedures, for specific conditions. It would make no sense to present the patient with the sort of technical evidence that would be taken into account by an expert in the field. It would be generally accepted that it is the duty of health professionals to provide the patient with as much understanding of these issues as possible and to engage them in the decision-making. Exactly the same applies in childhood.

Testing minors at a young age for late-onset genetic diseases denies them the option of deciding about testing later in life, although this would have been possible.

Second, professional recommendations[27] implicitly acknowledge that there is a substantial difference between (1) a predictive genetic test for adult-onset diseases and (2) a medical diagnosis that says something about the current medical condition of the patient or a vaccination with the aim of prolonging immunity to specific diseases. The impact of genetic testing differs often from other types of medical interventions. We do not say that genetic testing is unique in one of these respects, but the cumulative complexity of these elements requires that genetic testing should receive careful consideration. First, genetic information is familial. As a result, the test results of a person might have direct implications for relatives, including offspring. Second, presymptomatic and predictive genetic testing provides information about the medical future of a healthy person. The person who undergoes a predictive test is not (yet) a patient. In other words, it concerns a real presymptomatic diagnosis. Third, there is a concern that genetic testing may lead to misuse of the genetic information by third parties, including insurers and employers, to deny, limit, or cancel health insurance or to discriminate against the person in the workplace. Furthermore, genetic information may also lead to stigmatization and discrimination within the community. A genetic test or the result of a genetic test may also lead to psychological distress. Increased anxiety, distress, depression, changed familial relations, changes in self-image, behavioral changes, and an impact on life planning are some of the potential negative psychosocial outcomes that have been reported in the literature as a result of a genetic test. It is clear that these features are not valid for all genetic data and that some of these characteristics are valid also for nongenetic medical data. In this way, genetic information differs not so much principally but rather gradually from other medical data.[28]

Because of the specific character of genetic information and because minors are urged as much as possible to participate in decisions affecting their own health, we support the professional guidelines in their position of delaying such genetic testing. Predictive genetic testing, as well as medical diagnostic evaluations or treatment decisions, should be performed on children if they provide timely medical benefit to the children. If the medical benefits are absent or uncertain, the justification for a specific medical intervention is less compelling. In addition, there should be an effective treatment or intervention for patients identified through early detection, with evidence of early treatment leading to better outcomes than later treatment. For predictive genetic testing for adult-onset diseases, the advantages in terms of medical benefit are absent because of the long period until the onset of the disease and the absence (in the cases studied here) of therapeutic or preventive measures. Because minors do not have any prospect of benefitting from effective treatment now, the question is if nonmedical benefits might provide convincing arguments about whether testing should be undertaken.

NONMEDICAL BENEFITS AND HARM

Existing professional recommendations have always defended a cautionary position toward predictive genetic testing in minors because of the potential harm that it may lead to. For example, according to the Working Party of the Clinical Genetics Society,[10]

> The potential harms caused by childhood genetic testing might include damage to the child's self-esteem, distortion of the family's perceptions of the child, loss of future adult autonomy and confidentiality, discrimination against the child in education, employment or insurance, and adverse effects on the child's capacity to form future relationships.

It is also asserted that the burden of certainty–knowing that one will get a genetic disease later in life–may result in severe psychological harm and even suicide, suicide attempts, and psychiatric hospitalizations[29]; this may be termed the unbearability of certainty.[30] However, there are difficulties in extrapolating these studies to children. It is simply not known if at-risk individuals would be better or worse if they had the test. There is a lack of informative studies about the psychological impact of predictive genetic testing on minors. Only a few studies have been performed. Grosfeld *et al.*[31] studied the psychological reactions of 22 parental couples and 3 single parents of children tested for multiple endocrine neoplasia type 2. The study showed that parents with favorable test results in their children experienced significantly less anxiety, whereas parents with children having the mutation experienced moderate to high levels of anxiety. Testing for familial adenomatous polyposis (FAP) did not lead to clinically significant changes in psychological symptoms (eg, depression, anxiety behavior, and competence) in tested children or their parents 3 months after testing.[32] Michie *et al.*[33] also did not find significant increases in symptoms of anxiety or depression, but they did find significantly higher anxiety among children who tested positive for FAP versus their tested-negative counterparts. A recent study by Codori *et al.*[34] showed that children did not suffer clinically significant psychological distress after testing. However, the same study also showed that "it would be clinically irresponsible to ignore the increases in depression and anxiety among children with positive siblings." In general, however, the reactions to the genetic test were rather positive, and negative effects were mostly transitory. Of course, the data are too limited to make conclusive decisions regarding genetic testing in minors, and the favorable reactions may be attributable to the view of multiple endocrine neoplasia type 2 and FAP as treatable diseases.

The introduction of nonmedical benefits and harm in the arguments regarding predictive genetic testing for minors for adult-onset genetic diseases integrates a concept of best interests that involves far more complex matters than physical criteria alone. Rhodes' critique[5] of best interests, which measure benefit primarily in terms of physical gains, is in the line with the critique issued by the British Medical Association (BMA). The BMA[35] developed a range of general factors with social and emotional elements that need to be considered when a young patient's best interests are assessed in relation to treatment. Among other things, the BMA emphasizes "the patient's own ascertainable wishes, feelings and values," "the patient's ability to understand what is proposed and weigh up the alternatives," "the patient's potential to participate more in the decision, if provided with additional support or explanations," "the patient's physical and emotional needs," "the risk of harm or suffering for the patient," "the views of parents and family," and "the prioritizing of options that maximize the patient's future opportunities and choices." However, even on the basis of this broad interpretation of the concept of best interests, contrary to the conclusion of Rhodes, the BMA comes to the conclusion that in the case of predictive testing for late-onset disorders, "testing in children should be opposed because it would undermine the future adult's right to make his or her own decision about whether to be tested." Rhodes describes various factors that led her to conclude that "pediatricians and geneticists should encourage reluctant parents to have their children tested–and then accept the parents' decision."[5] First, Rhodes mentions that

> 50–70% of those tested would also get the good news that they are unaffected. I take those odds alone to be an overwhelming reason for testing children.

She also stresses that almost everyone who is tested can benefit, at least by being relieved of uncertainty. Second, she criticizes the concept of unbearability of certainty because deceiving or withholding information from children can be harmful. Third, she advances that this information can become part of identity: "When this information is disclosed later in life, it can be jarring to identity and very hard to internalize and accept." Fourth, she asserts that delays in learning a child's genetic status prolong the parents' uncertainty and anxiety. In what follows, we dissect these 4 arguments and discuss alternatives.

Good News

It is clear that for those individuals who do not carry the mutation for a particular disorder, it is possible that many adverse consequences from assumptions of being affected may be averted. There are indeed examples of such good news reported in the literature.[36] People who undergo genetic testing and receive "good news" may learn definitively, or with a high probability, that they will not develop the disease and transmit it to their children. For those individuals who have the faulty mutation for a particular disorder, it allows anticipation of the future and appropriate planning. The act of testing could also be a catalyst for frank discussion with the family on the genetic aspects of the disorder, which the child is already encountering by living within an affected family. It may prevent children from being deceived or barred from information about genetic diseases.

A major problem related to the argument that genetic testing of minors might bring "good news" is that it does not sufficiently consider the consequences for those who test positive. Not only does it remove the opportunity for the child to make a decision about having the genetic test later in life, but it also often introduces more uncertainty and other risks. It is difficult for the child to live in the knowledge that, while not sick now, he or she has a greater likelihood of becoming sick in the future. Depending on the disease, this predictive value of the results might vary enormously. Although a child with a positive test result for Huntington's disease is almost certain to develop the disorder, many women with a positive test result for breast cancer do not get that disease. Even when a test can detect a mutation capable of causing a disease, the test may not be able to predict with certainty whether disease symptoms will appear, when they will appear, or what the degree of severity will be. In addition, the child is at risk for such harm as stigmatization and discrimination, increased anxiety, distress, depression, changed familial relations, changes in self-image, behavioral changes, and an adverse impact on life planning.

In addition, even favorable news and reassurance may affect people's self-image and the family dynamics. It may lead to what is called "survivor guilt" and a sense of stigmatization and exclusion from affected members of the family.[37] The following quote from Hayes[38] shows that receiving good news may lead to psychological and social distress and troubling family relations:

> It was a long and complex process. Blood samples from numerous members of my family had to be collected and analyzed. I underwent several months of genetic counseling to determine my ability to cope with any possible outcome. After a period of months, nothing remained but the nerve-racking wait for the results. . . . Finally, the wait was over: my test was negative. The DNA analysis has shown with 96 percent certainty (later increased to 99 percent, with refinement of the testing process) that I had not inherited the gene for Huntington's disease. When I learned the results I cried and laughed. It took months for the news to sink in. I am still adjusting. . . . The incomparable relief I felt at finally being free of the fear and uncertainty was tempered by the painful knowledge that other family members had not been and would not be so lucky.

Moreover, the message of "good news" fluctuates between analytic sensitivity, the ability of a test to detect the various mutations that it was designed to detect, and clinical sensitivity, the ability of the test to detect all patients who will get the disease. Even if the test result is negative, patients might still retain a residual risk of developing the disease. The study of Grosfeld *et al.*[31] showed that even parents who received "good news" often were not reassured by the DNA test result, wanted confirmation of the test result, and were willing to continue follow-up of their children who did not have the faulty mutations.

Unbearability of Knowing

In her analysis, Rhodes[5] is distinguishing 2 separate options:

> Either the testing is not performed and the child is raised without the family and the child having information about whether or not the child has inherited the mutation, or the testing is performed and the family raises the child in the context of knowing that the child has or has not inherited the mutation. Logically there are no other options.

However, in this way the author does not take into account the intermediate option of informing children about their genetic risk without performing the predictive genetic test in childhood.

It is clear that it would be wrong to deceive or withhold relevant information from children. Genetic information is relevant and important to children, and concealment of this information denies to children important details about themselves. As Clayton stated, "long experience demonstrates that hiding information from children usually does not work and that efforts to keep secrets leave children feeling deceived and abandoned."[39] Withholding information from the child results in withholding "the power to make informed choices."[40] Studies show that "those who felt that information had been deliberately withheld from them were deeply resentful and expressed anger and feelings of disempowerment as a result."[40] Parents are responsible for the health and well-being of their children and have the important responsibility to communicate the genetic risk to their children and to support them in dealing with that risk. We underline the familial character of genetic information and emphasize that genetic risk information should be discussed as openly as possible in a family context. Telling children about genetic risk is generally seen as a parent's responsibility,[41] but health professionals may be needed in a supporting role.

Of course, open communication about genetic risk between parents and children is not always easy. Feelings of guilt and blame may affect family relationships and hinder communication about genetic risk.[42,43] Psychological defense mechanisms such as denial and rationalization have been observed as reasons for parents not telling children in Huntington's disease families.[44] In hereditary breast and ovarian cancer families, studies suggest that some patients use similar strategies to minimize the emotional impact of the risk to themselves and their daughters. As a result, there may be interruptions in the transmission of risk information to some children if a parent declines to tell[41] or if a sibling or parent dies.[45] This may be compounded by a tendency for aunts or uncles not to speak with nieces or nephews and/or a sense that they do not have the authority to do so.

Having open communication between parents and a child about genetic risk and the possibility of a genetic test as soon as the minor has the competence to provide consent to it seems to us a better alternative than just testing the minor in childhood. Testing children as an isolated act is not an alternative for open communication: the child might be told once, and it may not be discussed again. It might also confuse or frighten the child. Testing in childhood does not ensure further communication and follow-up. When a minor understands the information given to him and the implications of a positive and negative genetic test for him, his partner, and his possible offspring, he can make a decision (eventually together with his parents) about performing this test.

A major difference between informing children of the fact that they have been adopted or created with gamete donation and performing a genetic test is that the information is already there in the former case. It is clear that families should not be discouraged from telling their children that they have a genetic risk. Just as in the situation of adoption, informing children about the presence of this risk (or the fact of being adopted) allows children to ask questions and to make statements about undergoing a predictive genetic test themselves (or to learn more about the adoption). Our position is in line with existing guidelines that state that if parents discuss the "child's risk with the child, they will be able to explain to the child that testing will be available at the discretion of the child when the child reaches maturity."[13]

Identity and Adjustment

According to Rhodes,[5]

> Information becomes part of identity. When a child learns personal genetic information early in life, it can be absorbed and accommodated into identity. When the information is disclosed later in life, it can be jarring to identity and very hard to internalize and accept.

We clearly follow this idea of Rhodes, in the sense that information about genetic risk should not be withheld from children. This of course requires that the children understand "what they are being told, can make sense of the information (according to their age), are not fearful of it and are able to talk about it and ask questions whenever they like."[46] Children should be offered the opportunity to incorporate this information into their self-image, to look to the future, and to plan what to do about their genetic risk. As Skirton notes, "the advantage of telling a child gradually as the opportunity presents itself or as questions arise was stressed by a number of respondents, who felt that there were benefits in being able to absorb the information slowly and so gradually adjust to the altered potential of life."[40] Also, the Genetic Interest Group reports that "children can cope with information about themselves from an early age and that it is much more often the adult who has a problem in giving information."[47]

Disclosure of disease risk may also benefit children when they are told at a young age instead of being told as adolescents or adults. In her study, Skirton reported that "it is easier for a young person to deal with the news of risk than it would be at an older age."[40] From this perspective, we advance that in building up an identity, it is important that a child be told about the risk for a disease that exists within a family, but it is not necessarily important to tell him whether or not he carries the faulty mutation. We recommend that the child be made aware of genetic conditions in the family that may have an impact on him and that the child be informed about his genetic risk. This enables the child to make a free choice about genetic testing later in life.

Parental Anxiety and Uncertainty

Rhodes[5] also advances that delays in learning a child's genetic status prolong parents' uncertainty and anxiety. Various studies[48–50] have indeed shown that parents believe that detection in childhood might help prepare their children and themselves psychologically for the future. Various parents that have been tested themselves for a specific disease and know that their children are at risk find sometimes the uncertainty of not knowing more burdensome than receiving a negative or positive test result. Therefore, parents often conclude that they should be able to consent to genetic testing in their children for diseases that have their onset only in adulthood.

However, it should be clear that only 10% to 15% of at-risk adults opt for Huntington's disease genetic testing. Studies[37] that have explored the reasons for this low take-up have shown the most important reasons for not being tested were increased risk to children if one was found to be a gene carrier; absence of an effective cure, potential loss of health insurance, financial costs of testing, and the inability to "undo" the knowledge. The risk to relatives, lack of treatment or cure, and fear of losing one's health insurance are important factors in the decision not to be tested.

If parents would test their young children, there would not be able to make this personal evaluation. In this context, it is relevant to refer to the notion of the "right not to know," which is increasingly accepted also at the international level. The right not to know has been recognized, for example, in the European Convention for the Protection of Human Rights and Dignity of the Human Being with Regard to the Application of Biology and Medicine, which states [in article 10(2)] that

> Everyone is entitled to know any information collected about his or her health. However, the wishes of individuals not to be so informed shall be observed.

Similarly, the UNESCO Universal Declaration on the Human Genome and Human Rights states (in article 5c) that

> The right of every individual to decide whether or not to be informed of the results of genetic examination and the resulting consequences should be respected.

Testing a minor eliminates the possibility of the minor making use of this "right not to know."

As we have stated previously, it is also clear that testing does not take away all anxieties and uncertainties. Indeed, anxiety and distress might increase when children test positive. We suggest that the most important reasons for letting children decide later about performing such a genetic test probably concern those who will be affected by the disorder. Rhodes[5] refers in her article to the fact that because parents, "far more than health care providers or policy makers, are the ones who will be living with the repercussions of their choices, the judgment should be theirs." However, we would like to advance the opposite statement, that because minors, far more than their parents, will be living with the repercussions, they should be able to decide about participation in such a genetic test. The presence of severe anxieties and uncertainties in parents about a potential genetic mutation might be an indication for further psychological support in order to address these emotions rather than a clear indication for testing.

CONCLUSIONS

In her article,[5] Rhodes challenges the recommendations of the genetic and pediatric communities regarding predictive genetic testing for late-onset genetic disorders for which there are no preventive or therapeutic measures. In this article, we have entered into a dialogue and discussed her various arguments. At the end of this article, we would first like to emphasize that the main reason for the existing guideline not to perform such testing is the fact that testing does not provide "scope for treatment which to any essential degree prevents, defers or alleviates the outbreak of disease or the consequences of the outbreak of disease."[15] Second, we question if the parents are entitled to make such decisions. On the one hand, one can intuitively understand that parents feel responsible for the well-being of their children and want to take actions that benefit them and their offspring. On the other hand, we suggest that the responsibility of parents does not go so far that they should be responsible for deciding about performing such a test in childhood. This would be a moral hypertrophy of responsibility.[51] In this way, responsibility is being stretched endlessly.[52] We suggest, however, that parents are responsible for the health and well-being of their children and have the important responsibility of communicating the genetic risk to their children and supporting them in dealing with that risk. We underline the familial character of genetic information and emphasize that genetic risk information should be discussed openly in a family context. However, we also emphasize the individual character of this information, the individual decision to undergo a genetic test, and the individual choice to share the result of the genetic test with others. Furthermore, psychosocial concerns incite us to caution. Very few studies have been reported concerning the psychosocial consequences of predictive genetic testing in childhood or adolescence. Even if they did not report long-term negative psychological and social impact, there is no reason to assume that such testing in minors should not be assessed with caution. Studies with adults incite us to do so. Even if there is no clear evidence of harm, there is no clear evidence of benefit. Letting minors decide about predictive genetic testing for late-onset diseases means respecting them as humans with their own individuality and responsibility. Parents and genetic services should be supportive so that minors at risk are informed about their genetic risk and are able to make use of the opportunities of genetic testing later in life.

ACKNOWLEDGEMENT

The research for this contribution was supported by the Eurogentest Network of Excellence of the European Union (FP6–512148).

DISCLOSURES

There is nothing to report.

REFERENCES

1. Harper P, Clarke A. Genetics, society and clinical practice. Oxford, England: Bios; 1997.
2. Biesecker BB. Future directions in genetic counseling: practical and ethical considerations. *Kennedy Inst Ethics J* 1998; 8: 145–160.
3. Mahowald MB, Verp MS, Anderson RR. Genetic counseling: clinical and ethical challenges. *Annu Rev Genet* 1998; 32: 547–559.
4. Allan D. Ethical boundaries in genetic testing. *CMAJ* 1996; 154: 241–244.
5. Rhodes R. Why test children for adult-onset genetic diseases? *Mt Sinai J Med* 2006; 73: 609–616.
6. The American Society of Human Genetics Board of Directors and the American College of Medical Genetics Board of Directors. Points to consider: ethical,

legal, and psychological implications of genetic testing in children and adolescents. *Am J Hum Genet* 1995; 57: 1233–1241.

7. Human Genetics Society of Australasia. Child testing policy (2005). Available at: http://www.hgsa.com.au. Accessed March 2008.
8. German Society of Human Genetics. Statement on genetic diagnosis in children and adolescents (1995). Available at: http://www.gfhev.de/en/gfh. Accessed March 2008.
9. Committee on Bioethics. Ethical issues with genetic testing in pediatrics. *Pediatrics* 2001; 107: 1451.
10. Clarke A. The genetic testing of children. Working Party of the Clinical Genetics Society (UK). *J Med Genet* 1994; 31: 785–797.
11. Canadian College of Medical Geneticists. Genetic testing of children (2000). Available at: http://ccmg.medical.org. Accessed March 2008.
12. Bioethics Committee Canadian Paediatric Society. Guidelines for genetic testing of healthy children. *Paediatr Child Health* 2003; 8: 42–45.
13. American Medical Association. Testing children for genetic status (1995). Available at: http://www.ama-assn.org. Accessed March 2008.
14. Committee for Public Relations and Ethical Issues of the German Society of Human Genetics. Statement on genetic diagnosis in children and adolescents (1995). Available at: http://www.gfhev.de/en/gfh. Accessed March 2008.
15. Danish Council of Ethics. Genetic investigation of healthy subjects–report on presymptomatic gene diagnosis (2001). Available at: http://www.etiskraad.dk/sw293.asp. Accessed March 2008.
16. Human Genetics Society of Australia. DNA presymptomatic and predictive testing for genetic disorders (2002). Available at: http://www.hgsa.com.au/policy/dnaptgd.html. Accessed March 2008.
17. Belgian Society of Human Genetics. Guidelines for predictive genetic testing for late onset disorders (2003). Available at: http://www.beshg.be. Accessed March 2008.
18. European Society of Human Genetics. Provision of genetic services in Europe–current practices and issues (2001). Available at: http://www.eshg.org. Accessed March 2008.
19. Stultiens L, Goffin T, Borry P, et al. Minors and informed consent: a comparative approach. *Eur J Health Law* 2007; 14: 21–46.
20. Health Act–Law No. 546 of 24 June 2005.
21. Evers-Kiebooms G, Welkenhuysen M, Claes E, et al. The psychological complexity of predictive testing for late onset neurogenetic diseases and hereditary cancers: implications for multidisciplinary counselling and for genetic education. *Soc Sci Med* 2000; 51: 831–841.
22. Buchanan AE, Brock DW. Deciding for Others. Cambridge, England: Cambridge University Press; 1989.
23. Charter of Fundamental Rights of the European Union (2000). Available at: http://www.europarl.europa.eu/charter/default_en.htm. Accessed March 2008.
24. UNESCO. Universal Declaration on the Human Genome and Human Rights (1997). Available at: http://portal.unesco.org. Accessed March 2008.
25. Alderson P, Montgomery J. Health Care Choices: Making Decisions with Children. London, England: Institute for Public Policy Research; 1996.
26. Royal College of Psychiatrists Report. Appendix 2–Children's level of understanding of medical decisions. Guidance for researchers and for ethics committees on psychiatric research involving human participants (2001). Available at: http://www.rcpsych.ac.uk. Accessed March 2008.
27. Borry P, Stultiens L, Nys H, et al. Presymptomatic and predictive genetic testing in minors: a systematic review of guidelines and position papers. *Clin Genet* 2006; 70: 374–381.
28. Dierickx K. Genetic privacy and the family. *Ind J Fam Stud* 2004; 2: 68–80.
29. Bird TD. Outrageous fortune: the risk of suicide in genetic testing for Huntington disease. *Am J Hum Genet* 1999; 64: 1289–1292.
30. Almqvist EW, Bloch M, Brinkman R, et al. A worldwide assessment of the frequency of suicide, suicide attempts, or psychiatric hospitalization after predictive testing for Huntington disease. *Am J Hum Genet* 1999; 64: 1293–1304.
31. Grosfeld FJ, Beemer FA, Lips CJ, et al. Parents' responses to disclosure of genetic test results of their children. *Am J Med Genet* 2000; 94: 316–323.
32. Codori AM, Petersen GM, Boyd PA, et al. Genetic testing for cancer in children. Short-term psychological effect. *Arch Pediatr Adolesc Med* 1996; 150: 1131–1138.
33. Michie S, Bobrow M, Marteau TM. Predictive genetic testing in children and adults: a study of emotional impact. *J Med Genet* 2001; 38: 519.
34. Codori AM, Zawacki KL, Petersen GM, et al. Genetic testing for hereditary colorectal cancer in children: long-term psychological effects. *Am J Med Genet* 2003; 116A: 117–128.
35. British Medical Association. Human genetics: choice and responsibility. Oxford, England: Oxford University Press; 1998.
36. Fanos JH. Developmental tasks of childhood and adolescence: implications for genetic testing. *Am J Med Genet* 1997; 71: 22–28.
37. Quaid KA, Morris M. Reluctance to undergo predictive testing: the case of Huntington disease. *Am J Med Genet* 1993; 45: 41–45.
38. Hayes CV. Genetic testing for Huntington's disease–a family issue. *N Engl J Med* 1992; 327: 1449–1451.
39. Clayton EW. Genetic testing in children. *J Med Philos* 1997; 22: 233–251.
40. Skirton H. Telling the children. In: Clarke A, editor. The Genetic Testing of Children. Oxford, England: Bios; 1998; 103–111.
41. Forrest K, Simpson SA, Wilson BJ, et al. To tell or not to tell: barriers and facilitators in family communication about genetic risk. *Clin Genet* 2003; 64: 317–326.
42. Speice J, McDaniel SH, Rowley PT, Loader S. Family issues in a psychoeducation group for women with a BRCA mutation. *Clin Genet* 2002; 62: 121–127.
43. Fanos JH, Johnson JP. Barriers to carrier testing for adult cystic fibrosis sibs: the importance of not knowing. *Am J Med Genet* 1995; 59: 85–91.
44. Martindale B. Huntington's chorea: some psychodynamics seen in those at risk and in the responses of the helping professions. *Br J Psychiatry* 1987; 150: 319–323.

45. Peterson SK, Watts BG, Koehly LM, et al. How families communicate about HNPCC genetic testing: findings from a qualitative study. *Am J Med Genet* 2003; 119C: 78–86.
46. Malpas PJ. Why tell asymptomatic children of the risk of an adult-onset disease in the family but not test them for it? *J Med Ethics* 2006; 32: 639–642.
47. Dalby S. GIG response to the UK Clinical Genetics Society report "The genetic testing of children". *J Med Genet* 1995; 32: 490–491.
48. Campbell E, Ross LF. Parental attitudes and beliefs regarding the genetic testing of children. *Community Genet* 2005; 8: 94–102.
49. Hamann HA, Croyle RT, Venne VL, et al. Attitudes toward the genetic testing of children among adults in a Utah-based kindred tested for a BRCA1 mutation. *Am J Med Genet* 2000; 92: 25–32.
50. Benkendorf JL, Reutenauer JE, Hughes CA, et al. Patients' attitudes about autonomy and confidentiality in genetic testing for breast-ovarian cancer susceptibility. *Am J Med Genet* 1997; 73: 296–303.
51. ten Have HA. Responsibilities and rationalities: should the patient be blamed? Paper one: The distinction between prospective and retrospective responsibility. *Health Care Anal* 1994; 2: 119–123.
52. Dierickx K. Autonomous decisions in assisted procreation and genetic diagnosis. In: Haker H, Beyleveld D, editors. The Ethics of Genetics in Human Procreation. Aldershot, England: Ashgate; 2000; 131–138.

NEW DRUGS AND THE FDA

Chemokine (C-C Motif) Receptor 5 Entry Inhibitors: New Class of Drugs for Acquired Immune Deficiency Syndrome

Andrew Jacovina, PhD,[1] and Maria New, MD[1,2]

[1]Department of Genetics and Genomic Science, Mount Sinai School of Medicine, New York, NY and
[2]Departments of Pediatrics, Mount Sinai School of Medicine, New York, NY

Recently, a new anti–acquired immune deficiency syndrome (AIDS) drug called maraviroc was shown to be effective against human immunodeficiency virus (HIV) in patients who did not respond well to other antiretroviral therapies.[1,2] Maraviroc represents a novel class of drugs called chemokine (C-C motif) receptor 5 (CCR5) entry inhibitors. Unlike other antiretrovirals currently in use, it targets a human cell surface protein to which HIV attaches. If it is approved by the Food and Drug Administration, maraviroc would be the first drug against AIDS that targets the human host cell rather than the virus itself.

HIV infection commences with the binding of viral envelope proteins to CD-4 receptors on a host immune cell. Although CD-4 binding is required, it is not sufficient in itself for HIV infection. In order to gain entry into the cell, HIV must also bind a host coreceptor, either CCR5 or chemokine (C-X-C motif) receptor 4 (CXCR4). Interactions between HIV envelope proteins and host cell receptors trigger the fusion of the viral lipid membrane with that of the host cell.[3]

The human CCR5 gene encodes a 352 amino acid (relative molecular mass = 40,600 Da) G protein–coupled chemokine receptor that plays a role in the inflammatory response. It binds several ligands, including RANTES (regulated on activation, normal T expressed and secreted), macrophage inflammatory protein 1α, and macrophage inflammatory protein 1β.[3,4] CCR5 regulates the development, recruitment, and activation of T cells and phagocytic cells by conveying extracellular chemokine-mediated signals to the nucleus via the mitogen-activated protein kinase and c-Jun N-terminal kinase signal transduction pathways.[4]

In 1996, several studies described rare individuals who exhibited a resistance to HIV infection despite multiple exposures to the virus.[5,6] These individuals possess a mutated form of the CCRS gene dubbed Δ*ccr5*. The mutation consists of a 32-bp deletion that results in a truncated, functionless protein.[4] Homozygous Δ*ccr5* individuals do not express any CCR5 and are largely resistant to infection by HIV-1.[5,6] Heterozygous Δ*ccr5* individuals exhibit elevated HIV resistance, but if they do become infected, they have reduced viral loads, and the progression of their disease is generally delayed by 2 to 3 years.[5,6] The Δ*ccr5* allele appears to have little negative impact on human immune status or general health. Studies in CCR5-deficient mice have revealed only limited defects in the cell-mediated immune response.[4] It is thought that critical functions of the missing CCR5 may be carried out by other chemokine receptors and their ligands because of the intrinsic redundancy of the chemokine system.[5,6]

The Δ*ccr5* allele is found in Caucasian Europeans and to a lesser degree in Middle Eastern and Indian populations, but it is not detected in Africa or the rest of Asia. Within Europe, the Δ*ccr5* allele displays a distinct geographical frequency distribution with a northeast to southwest downhill cline. It occurs in 15% of the population of the Baltic region but is present in only 4% of Sardinians. Overall, the allele

Address Correspondence to:

Maria New
Department of Genetics and
Genomic Science
Mount Sinai School of Medicine
New York, NY
Email: maria.new@mssm.edu

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20050

© 2008 Mount Sinai School of Medicine

is found in 10% of the Caucasian European population, of which 1% is homozygous for the mutation.[7,8] Genetic and historical evidence suggests that Viking raiders may have been instrumental in spreading the Δ*ccr5* allele from its Scandinavian origin to the rest of Europe. This hypothesis is supported by the high frequency of Δ*ccr5* in regions such as Brittany (13.5%) and Estonia (13.3%), in which Viking activity was common.

The Δ*ccr5* mutation arose within the last 2250 years.[7] The high frequency of its occurrence in such a relatively short period of time suggests that a strong selective factor may have been involved in the persistence and spread of the Δ*ccr5* allele.[9] Because Δ*ccr5* arose before the introduction of HIV-1 into Europe, other pathogens, including bubonic plague, dysentery, and smallpox, have been implicated as the responsible selective agent. The most likely candidate is smallpox, a viral pathogen that has continually exerted strong selective pressure on human populations until its global eradication in the 1970s. Because poxviruses also infect lymphocytes using chemokine receptors,[10] it is biologically plausible that a disruptive mutation in the CCR5 chemokine receptor could confer resistance against both smallpox and HIV.[9]

Maraviroc is an orally administered small-molecule CCR5 antagonist that has demonstrated a promising efficacy and safety profile in phase 2a clinical trials.[11] Its novel mode of action prevents HIV from entering uninfected cells by blocking its predominant entry route, the CCR5 coreceptor.[2] Pfizer released the results of its maraviroc phase 2b/3 clinical trials at the 14th Conference on Retroviruses and Opportunistic Infections (February 25–28, 2007).[2] In double-blind, randomized trials, 1049 HIV-positive patients received 3 to 6 antiretrovirals combined with either a placebo or maraviroc dosed once or twice daily. A 24-week analysis showed that approximately twice as many patients receiving maraviroc achieved undetectable virus in the blood (<50 copies/mL of HIV RNA) in comparison with the control group. In addition, patients receiving maraviroc saw an increase in CD-4+ cell counts nearly twice that seen in the controls. The groups receiving maraviroc had a low rate of discontinuation due to adverse events that was similar to the rate of the group receiving standard antiretrovirals alone.[2] One concern regarding CCR5 antagonists is the possibility that selective pressure from these drugs might result in the emergence of more aggressive CXCR4-tropic HIV strains. However, maraviroc did not select for CXCR4-tropic variants after prolonged in vitro serial passage, and this suggests that maraviroc resistance may not be easy to acquire.[11]

Although there are more than 20 AIDS drugs on the market, unpleasant side effects, complex dosing regiments, and the emergence of drug resistance add to the urgent need for new classes of anti-AIDS drugs.[1] If approved, maraviroc would help meet that need because it represents the first new oral class of HIV drugs to be developed in more than a decade.

DISCLOSURES

There is nothing to report.

REFERENCES

1. Altman LK, Pollack A. Two new drugs offer options to fight H.I.V. in novel ways. *New York Times*. February 28, 2007; A1.
2. Lalezari J, Goodrich J, DeJesus E, et al. Efficacy and safety of maraviroc plus optimized background therapy in viremic ART-experienced patients infected with CCR5-tropic HIV-1: 24-week results of a phase 2b/3 study in the US and Canada. Paper presented at: 14th Conference on Retroviruses and Opportunistic Infections; February 25–28, 2007; Los Angeles, CA.
3. Moore JP, Doms RW. The entry of entry inhibitors: a fusion of science and medicine. *Proc Natl Acad Sci* 2003; 100: 10598–10602.
4. Mueller A, Strange P. The chemokine receptor, CCR5. *Int J Biochem Cell Biol* 2004; 36: 35–38.
5. Liu R, Paxton WA, Choe C, et al. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* 1996; 86: 367–377.
6. Paxton WA, Martin SR, Tse D, et al. Relative resistance to HIV-1 infection of CD4 lymphocytes from persons who remain uninfected despite multiple high-risk sexual exposure. *Nat Med* 1996; 2: 412–417.
7. Libert F, Cochaux P, Beckman G, et al. The Δccr5 mutation conferring protection against HIV-1 in Caucasian populations has a single and recent origin in Northeastern Europe. *Hum Mol Genet* 1998; 7: 399–406.
8. Lucotte G, Dieterlen F. More about the Viking hypothesis of origin of the Δ32 mutation in the CCR5 gene conferring resistance to HIV-1 infection. *Infect Genet Evol* 2003; 3: 293–295.
9. Galvani AP, Novembre J. The evolutionary history of the *CCR5-Δ32* HIV-resistance mutation. *Microbes Infect* 2005; 7: 302–309.
10. Lalani AS, Masters J, Zeng W, et al. Use of chemokine receptors by poxviruses. *Science* 1999; 286: 1968–1971.
11. Westby M, Smith-Burchnell C, Mori J, et al. Reduced maximal inhibition in phenotypic susceptibility assays indicates that viral strains resistant to the CCR5 antagonist maraviroc utilize inhibitor-bound receptor for entry. *J Virol* 2007; 81: 2359–2371.

Evaluation and Treatment of Thyroid Nodules: A Clinical Guide

Juan C. Galofré, MD, PhD,[1,2] Stylianos Lomvardias, MD,[3] and
Terry F. Davies, MD, FRCP[1]

[1] Thyroid Research Unit, Mount Sinai School of Medicine, James J. Peters VA Medical Center, New York, NY
[2] Departamento de Endocrinología y Nutrición, Clínica Universitaria, University of Navarra, Pamplona, Spain
[3] Department of Pathology, James J. Peters VA Medical Center, New York, NY

ABSTRACT

Thyroid nodules are inexplicably frequent and affect approximately one-third of the adult population. The appropriate clinical management is focused primarily on excluding thyroid cancer and also on evaluating thyroid dysfunction and mechanical obstruction. There remains no evidence that a benign thyroid nodule, once diagnosed appropriately, will progress to a malignant lesion. The initial evaluation should include a complete clinical review, a thyroid sonogram by an experienced sonographer, a laboratory assessment of thyroid function, and, where indicated, a cytological assessment of the nodule(s) by fine needle aspiration under ultrasound guidance. Only patients with suppressed serum thyroid-stimulating hormone levels, indicating hyperthyroidism, may need further evaluation by radioactive iodine uptake and scanning. Optimal treatment depends on the patient as well as the nodule characteristics. The usual options remain a simple annual follow-up to detect changes in nodule size and thyroid function and surgical removal. Levothyroxine therapy is now seldom indicated because of poor efficacy in nodule suppression and its inability to differentiate benign lesions from thyroid carcinoma. Clinical guidelines have a very arbitrary recommendation of aspiration biopsy in all lesions greater than 1 cm in size, but this proposal has no scientific basis and should always be viewed in the clinical context. ***Mt Sinai J Med 75:299–311, 2008.*** © *2008 Mount Sinai School of Medicine*

Address Correspondence to:

Terry F. Davies
Thyroid Research Unit
Mount Sinai Medical Center
Email: terry.davies@mssm.edu

Key Words: benign, diagnosis, malignant, management, multinodular, nodules, thyroid, treatment.

Thyroid nodules (TNs) are among the leading causes of referring patients to an endocrinologist. Indeed, by the time we reach 80 years of age, almost 80% of us have 1 or more such nodules.[1] Hence, the neck check has been introduced (Figure 1).[2] Usually in clinical practice, when a patient notices a lump in his neck, his principal concern is the possibility of malignancy. However, the vast majority of such lesions are benign, and the patient simply requires periodic follow-up after a careful study of his nodule's characteristics. Many clinicians are used to considering such nodules as hot or cold, but this type of thinking no longer applies. First, all nodules take up radioiodine, and rarely are they truly cold; second, the activity of a nodule or nodules is better assessed by the patient's serum thyroid-stimulating hormone (TSH) level, which is exquisitely sensitive to thyroid hormone output.

TNs may be single, multiple, cystic, or solid and may or may not be functional. Furthermore, they may be found in the presence of unrelated thyroid dysfunction (eg, secondary to inflammation). Accordingly, both the thyroid functional status and pathological evaluation need to be assessed. In addition, there is a wide spectrum of TNs, ranging from small incidentalomas to bothersome, symptomatic, large neck masses.

NOMENCLATURE

A wide variety of terms are used in discussions of TNs. *Benign* and *malignant* are self-explanatory. *Hot* refers to positive uptake of radioiodine on scanning, whereas *cold* is the old term suggesting a reduced uptake of radioiodine. Some people even use the

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20040

© 2008 Mount Sinai School of Medicine. † This article is a US government work and, as such, is in the public domain in the United States of America.

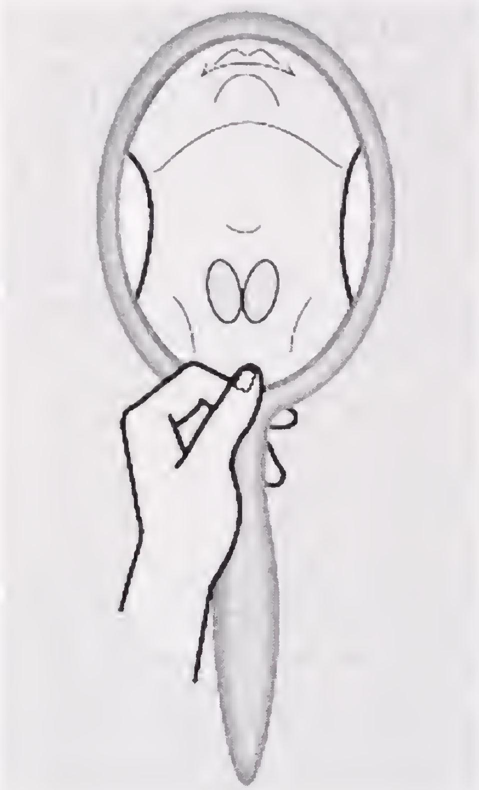


Fig 1. The thyroid neck check. All you need is a glass of water and a handheld mirror. Hold the mirror, focusing on your neck just between the Adam's apple and the collarbone, where your thyroid is located. While focusing on this area, tip your head back, take a drink of water, and swallow. As you swallow, look at your neck. Check for any bulges or enlargement in this area while you swallow. Do not confuse the Adam's apple with the thyroid gland. Reprinted with permission from http://www.thyroidawareness.com.[2] Copyright 2008, American Association of Clinical Endocrinologists. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.].

term *warm* for an in-between result. *Cystic* refers to the accumulation of fluid seen on ultrasound (US), whereas *complex* refers to a lesion with both solid and cystic components. *Multinodular goiter* is usually used when there are more than 2 nodules, whereas the term *colloid goiter* is now used by pathologists only to express the presence of colloid formation in a benign follicular lesion.

EPIDEMIOLOGY

The prevalence of TNs by physical examination is around 5% to 10% of the population, but physical examination of the thyroid gland has proven to be remarkably inaccurate with the introduction of thyroid sonography. This has caused most thyroidologists to perform their own thyroid sonographic examinations, and some of us believe that a real-time US examination is an essential part of all thyroid patient evaluations.[3] Indeed, US can discover TNs in 20% to 60% of the population, TNs being more frequent in women and the elderly[4–6] (Figure 2). The old fashioned evaluation by radioiodine scanning showed that 5% to 10% of palpated TNs were hot, whereas approximately 85% were cold, but data from US series show a much lower frequency of active nodules.[1] Nevertheless, a significant proportion of patients with multinodular thyroid disease have suppressed TSH levels and active nodules. The active single nodule, however, remains relatively uncommon.

FREQUENCY OF MALIGNANCY

The risk of a TN being thyroid cancer is between 1% and 10%.[7] It was once thought to be better to have more than 1 TN, but the prevalence of cancer is similar between solitary and multinodular goiters.[8] Usually, cancers take up less radioiodine than normal thyroid tissue and are called cold nodules on scintigraphy, but this does not mean that *cold nodule* is synonymous with *malignancy* because the vast majority of cold nodules are benign. Hence, routine thyroid scanning is a quite useless and expensive evaluation.

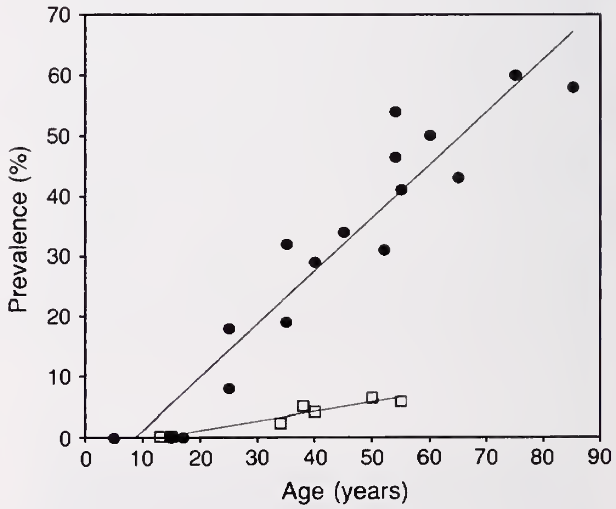


Fig 2. Prevalence of palpable thyroid nodules detected at autopsy or by ultrasonography (solid circles) or by palpation (open squares) in subjects without radiation exposure or known thyroid disease. Reprinted with permission from the *New England Journal of Medicine*.[4] Copyright 1993, Massachusetts Medical Society.

ETIOLOGY AND NATURAL HISTORY

Many environmental and constitutional factors have been related to the increased risk of developing TNs[9] (Table 1).

The first step in the transformation of the thyroid epithelial cell from the normal state to an abnormal state follows the generation of cellular hyperplasia. This may be induced by random somatic mutations or rearrangements in genes that encode mitogen-activated protein kinase pathway effectors,[10] the TSH receptor,[11,12] or a G-protein,[13] or this may follow a goitrogenic stimulus such as excessive TSH stimulation (Figure 3). This induction of growth exacerbates cellular mutagenesis with the formation of mutated cellular clones that generate the nodules.[13–15] When the mutation results in an active TSH receptor, the resulting nodule is autonomous, and the patient may develop hyperthyroidism and a suppressed TSH.

The evolution and natural history of TNs are variable and unpredictable. Some TNs will continue growing, and some will grow to a great extent and cause tracheal obstruction, whereas others will diminish with time, perhaps following infarction and hemorrhage. Hence, solid nodules tend to persist, and cystic lesions sometimes spontaneously disappear. As more nodules develop, the chance of an autonomous nodule being among them increases, and so a toxic multinodular goiter develops, most commonly in the elderly. Such TN activity may vary in time, and some autonomous lesions may evolve to cause changes in total thyroid function only after exposure to iodine; this is seen commonly after computed tomography (CT) scanning with iodine-containing contrast material. The iodine load allows the production of thyroid hormones more efficiently than before. Such development of autonomous nodules has been reported to occur in approximately 5% of cases.[16]

Table 1. *Factors Related to Increased Risk of Thyroid Nodule Development.*

Origin	Event
Environmental	Radiation exposure
	Smoking
	Iodine deficiency
	Drugs
	Emotional stress
	Infection
	Pregnancy
Constitutional	Family history
	Female gender
	Age

NOTE: All these events have been related to the appearance of thyroid nodules. Nevertheless, the specific role of these factors is poorly understood.

CLINICAL MANAGEMENT

In the study of TNs, clinicians must always be aware of the possibility of thyroid cancer, and certain clinical, analytical, US, and cytological data raise the suspicion of malignancy, but on their own, they may not prove it. However, the presence of these risk factors indicates a higher chance of malignancy.

Recent American and European guidelines recommend performing a complete evaluation only of those nodules that are larger than 1 cm.[5,17] However, the rationale for this recommendation is bewildering.[18,19] A malignant nodule starts its evolution as a single malignant cell. In theory, some of these malignant cells may have evolved from a previously benign abnormal cell within a TN, but the malignant transformation may also have started from a normal thyroid epithelial cell or from a thyroid stem cell.[20,21] Therefore, the malignant TNs will have been smaller than 1 cm, and a small nodule does not guarantee benignity or low risk.[22] In fact, some studies have calculated that the risk of malignancy in TNs less than 1 cm in size is as much as 10%. Of even more concern is the finding that 70% of these patients with small malignant nodules may have positive local lymph nodes.[7]

Office Evaluation

History

Normally functioning TNs are usually asymptomatic and are generally detected by routine neck examination (Figure 4). At presentation, the patient may complain of the presence of a slowly growing lump in the neck. TNs are rarely painful unless there has been an acute hemorrhage. Occasionally, TNs produce difficulty in swallowing that may indicate rapid growth, which is characteristic of malignancy. More often, it is vague discomfort on swallowing. The existence of a family history of benign TNs is common, and familial thyroid cancer is well described.[23] A moderate risk factor for malignancy is an age below 20 years and greater than 45 years, and male gender is also a factor,[5] whereas a highly suspicious indicator of malignancy includes a history of head and neck radiation.

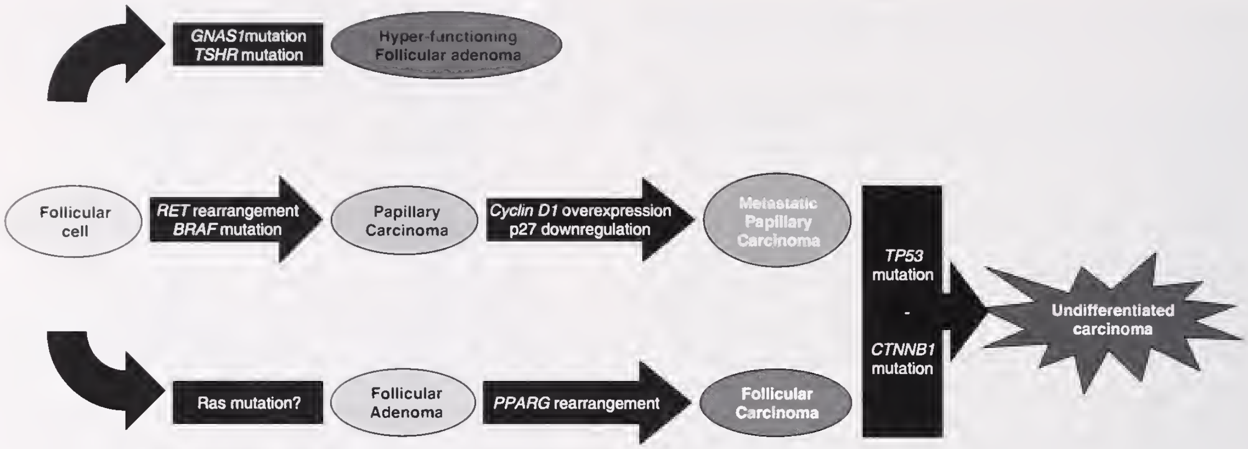


Fig 3. Proposed model of pathogenetic steps in thyroid follicular-cell neoplasia. Three different pathways have been proposed for neoplastic proliferation of thyroid follicular cells, including hyperfunctioning follicular thyroid adenoma (benign), papillary thyroid carcinoma, and follicular thyroid carcinoma. Activating mutations of *TSHR* and *GNAS1* occur in hyperfunctioning adenomas, whereas these are rare in malignancy. Genetic defects that result in activation of *RET* or *BRAF* represent an early event in the pathogenesis of papillary cancer. Underexpression of the cyclin-dependent kinase inhibitor p27 and overexpression of cyclin D1 may be predictors of the presence of lymph-node metastases. *Ras* proto-oncogene mutations have been related to the development of follicular thyroid adenomas. Rearrangements in *PPARG* have also been identified in follicular thyroid carcinoma. Undifferentiated malignancy is thought to evolve from both types of cancer, papillary and follicular, although different additional events such as inactivation of *p53* or *CTNNB1* mutation must also occur. **Abbreviations:** *BRAF*, V-raf murine sarcoma viral oncogene homolog B1; *CTNNB1*, catenin (cadherin-associated protein) beta 1; *GNAS1*, guanine nucleotide-binding α -subunit 1; *PPARG*, peroxisome proliferation-activated receptor- γ *RET*, rearranged during transfection; *TP53*, tumor protein 53; *TSHR*, thyroid-stimulating hormone receptor. Adapted from *Nature Reviews Cancer*.[10] [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.].

Examination

Observations of importance include several physical signs, rapid tumor growth (although thyroid cancers usually grow slowly), regional lymphadenopathy, and suspicion of distant metastases. If none of these factors are present, the clinical suspicion for malignancy is low (Table 2).

Laboratory Examination

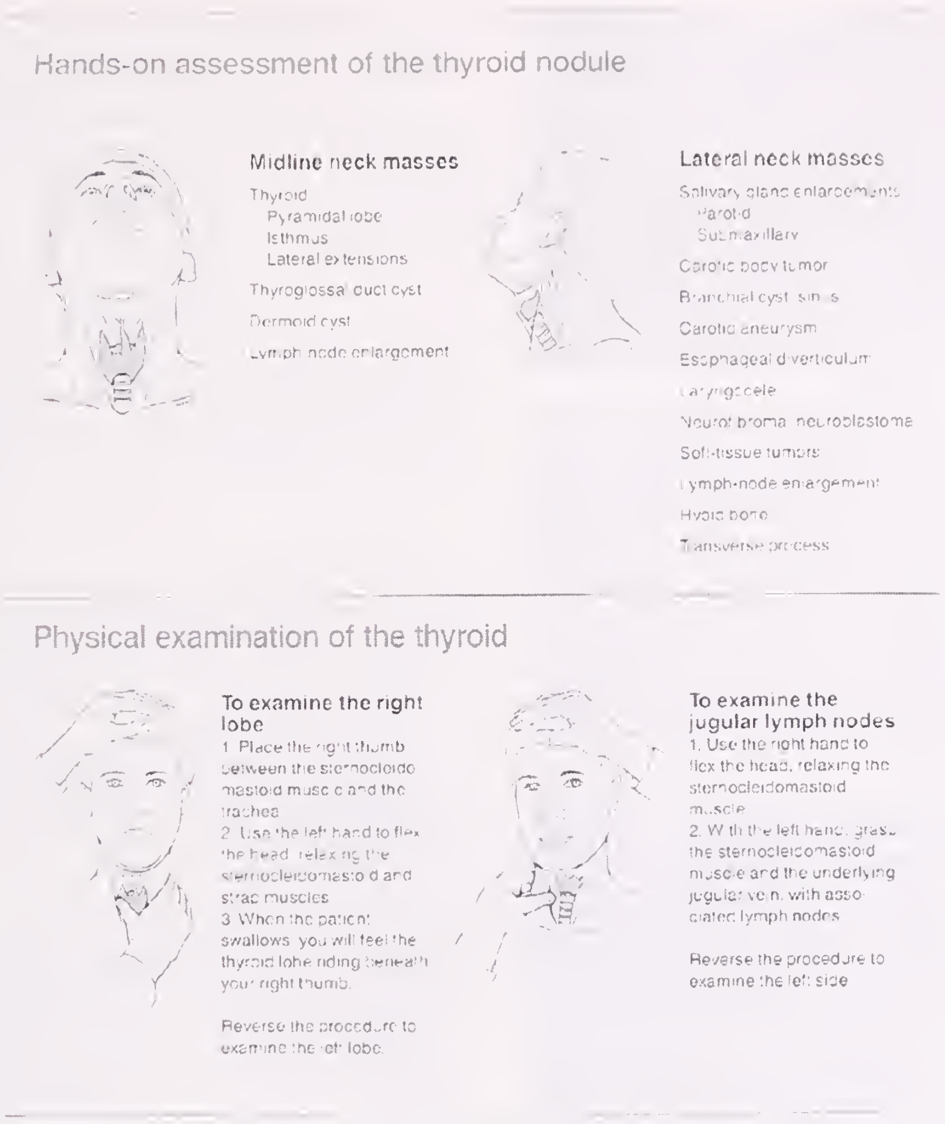
Thyroid Function Tests

The first study that should be performed in any patient with TNs is the determination of serum TSH to distinguish the autonomous nodules from the nonautonomous nodules.[5,16] A low serum TSH level usually means hyperthyroidism secondary to an autonomous nodule, and the patient should be managed accordingly.[24] In this setting, the next step is to obtain a sonogram and a radioiodine thyroid scan. If the thyroid scan shows that the TN concentrates a great deal of radioactivity, then no additional cytologic evaluation is necessary because functioning TNs are very rarely malignant.[5] However, it is also possible that the cause of the increased thyroid activity is not the nodule or nodules but rather coincidental Graves' disease. Graves' disease may present with multinodularity and can be confirmed by the presence of serum TSH receptor antibodies.[25] There is no evidence that Graves' disease itself leads to multinodularity, but such a presentation is more likely the reflection of 2 common disorders. The sonogram may also reveal that the palpated nodule and the active nodule are not coincidental.

TSH levels in TN patients are usually within the normal range. A high TSH indicates hypothyroidism, and measurement of thyroid antibodies is advisable in this situation to confirm coincidental Hashimoto's thyroiditis. There is an increased risk of thyroid cancer in patients with autoimmune thyroid disease. Interestingly, serum TSH concentrations may have an auxiliary value in raising suspicion of malignancy, as suggested by a recent study in which the risk of malignancy in a TN increased proportionately with serum TSH levels.[26]

Tumor Markers

Thyroglobulin is one of the most useful tumor markers for thyroid cancer recurrence once a patient has had his thyroid removed and any remaining tissue ablated with radioiodine. However, the determination of serum thyroglobulin has no value in the initial study of patients with TNs. All thyroid diseases may cause an increased serum thyroglobulin so that it



Hands-on assessment of the thyroid nodule

Midline neck masses

Thyroid
  Pyramidal lobe
  Isthmus
  Lateral extensions

Thyroglossal duct cyst

Dermoid cyst

Lymph node enlargement

Lateral neck masses

Salivary gland enlargements
  Parotid
  Submaxillary

Carotid body tumor

Branchial cyst sinus

Carotid aneurysm

Esophageal diverticulum

Laryngocele

Neurofibroma neuroblastoma

Soft-tissue tumors

Lymph-node enlargement

Hyoid bone

Transverse process

Physical examination of the thyroid

To examine the right lobe

1 Place the right thumb between the sternocleido mastoid muscle and the trachea
2 Use the left hand to flex the head, relaxing the sternocleidomastoid and strap muscles
3 When the patient swallows, you will feel the thyroid lobe riding beneath your right thumb.

Reverse the procedure to examine the left lobe.

To examine the jugular lymph nodes

1. Use the right hand to flex the head, relaxing the sternocleidomastoid muscle
2. With the left hand, grasp the sternocleidomastoid muscle and the underlying jugular vein, with associated lymph nodes

Reverse the procedure to examine the left side

Fig 4. Physical examination of the thyroid includes neck inspection and palpation of both thyroid lobes and lateral cervical lymph nodes. Reprinted with permission from *Acute Care Medicine.*[62] Copyright 1985. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.].

has no value as a routine thyroid function test. In contrast, a controversy exists about the usefulness of determining plasma calcitonin, oversecreted by medullary thyroid cancer cells, as part of the routine evaluation of TNs. The prevalence of medullary thyroid cancer in TNs ranges from 0.3% to 1.3%,[27] and this makes such testing highly cost-ineffective. Several European prospective studies of unselected TNs have demonstrated that regular measurement of serum calcitonin allows the detection of unsuspected medullary thyroid carcinoma with a frequency of 1 in 200 to 300 TNs, which gives a better sensitivity than fine needle aspiration (FNA).[27,28] Therefore, some authors have recommended routine calcitonin screening to improve the TN outcome.[28] However, in addition to the cost, the sensitivity and specificity of this practice have not been clearly determined.

Thyroid Autoantibodies

The measurement of thyroid autoantibodies is not needed in euthyroid patients with TNs but is recommended in the presence of thyroid dysfunction.[16] It is worth noting, however, that the presence of antibodies to thyroid peroxidase and thyroglobulin, which correlate with an intrathyroidal lymphocytic reaction, may confer an improved prognosis in patients with thyroid cancer.[29,30]

Table 2. *Indicators of Possible Malignancy.*

Clinical Indicators	Physical Signs	Sonographic Indicators
Past family history of thyroid cancer (especially in medullary cancer)	Hard consistency	Intranodular microcalcification
History of head and neck radiation	Fixation to surrounding structures	Intranodular hypervascularity (evaluated by Doppler)
Rapid tumor growth	Hoarseness (vocal cord palsy)	Hypoechogenicity
	Regional lymphadenopathy	Nodule with irregular borders
	Compressive symptoms (dysphagia, dyspnea, or coughing)	Absence of peripheral halo
	Large nodules (>4 cm)	Presence of regional lymphadenopathy

US Examination

Initial Evaluation

Thyroid US is the most accurate imaging technique for the detection of TNs, and this procedure is mandatory when a nodule is discovered at palpation or by any other method [eg, CT or magnetic resonance (MRI)]. It is important to know the nodule size, the number of nodules (presence of additional nonpalpable nodules), and their location and characteristics. In addition, US exploration is useful for detection of suspicious lymph nodes, which is important in the assessment of malignancy. Furthermore, US has the advantage of being a well-tolerated innocuous technique with low cost and no irradiation of the patient. The main drawbacks of thyroid US are the dependence on the operator's experience and the difficulty in viewing low retrosternal goiters, which may require CT/MRI evaluation.

Diagnostic Accuracy

The sonogram can help not only in the diagnosis of TNs but also in the evaluation of the nodules. However, US features of malignancy are not highly reliable when applied to the individual patient and simply add to the constellation of risk factors available for the clinician to consider. US characteristics suggestive of malignancy are summarized in Table 2. A combination of these findings may be highly predictive for malignancy. Although such sonographic characteristics are superior to just nodule size for identifying malignancy, a rapid increase in size is also an important risk factor.[31]

Other Imaging Techniques

Scintigraphy

This has been the imaging technique of choice to evaluate the functional activity of the thyroid gland and the TN in hyperthyroid states. As already mentioned, routine thyroid scintigraphy with radioactive iodine (RAI) or technetium is not helpful in the evaluation of euthyroid or hypothyroid patients with TNs. Scintigraphy uses ionized radiation; hence, it is contraindicated in pregnancy and lactation.

CT/MRI Scanning

These techniques are particularly useful in the evaluation of retrosternal goiters. They cannot distinguish benign TNs from malignant ones and are not as helpful as US in distinguishing benign lymph nodes from malignant ones. In TN patients, it is better to use MRI than CT because the high iodine load in the CT contrast medium can precipitate a toxic nodular goiter.[32] This is not a problem with MRI contrast media.

Positron Emission Tomography

Positron emission tomography evaluations are highly informative for thyroid functionality but have no advantage over other scanning procedures in the routine evaluation of TNs. A recent study showed that fluorodeoxyglucose metabolic activity did not significantly differ between malignant and benign follicular neoplasias.[33]

FNA

Indications

The general recommendation is that nodules greater than 1 cm in size should be biopsied. This of course makes no biological sense because even carcinomas start small. However, as a practical approach, this may be appropriate, but any fast growing lesion and any lesion with suspicious US characteristics should be aspirated. Patients with suppressed TSH levels should have RAI scanning beforehand because the biopsy of active nodules is not necessary.

Technique

A #27 or #25 needle, mounted in a 10-mL syringe, is passed through the lesion under US control. Negative pressure is induced and released before withdrawal of the needle so that the aspirate will remain in the needle hub. The aspirate is smeared on a glass slide, fixed, and stained for examination. Complications are rare. The patient may have at the very most some mild discomfort, swelling, and ecchymoses and most commonly has none of these.

Improving the Yield of FNA

US is an excellent aid for improving the accuracy of FNA cytology. US guidance can reduce the rate of inadequate smears from 15% to 3%.[34] Since 2007, all FNA procedures should have been performed with US guidance.[5]

Cytology

For determining whether a TN is malignant or benign, cytological diagnosis is, despite its limitations, still the most accurate and powerful preoperative study and must be included in the initial evaluation of all TNs except the autonomous variety.[5] Unfortunately, cytological reports are frequently confusing.[35] Cytological samples should be first divided into adequate and inadequate samples.[36] The term *nondiagnostic* is ambiguous and should be avoided in cytological reports. Some authors consider adequate only those specimens that include 6 or more groups of 10 to 20 well-preserved follicular cells on at least 2 slides.[34] Inadequate samples often necessitate a repeat FNA. Data from Mount Sinai in a series of almost 500 cytologic samples showed 43% of such samples to be inadequate specimens. This high percentage is related to the large number of different referring physicians (>50) who perform too few procedures to develop good technique and confirms the importance of a certain degree of skill and experience needed to yield useful data.[37]

Five Categories of Adequate Samples[36]

Benign This category includes adenomatous nodules, multinodular goiters, cysts, and thyroiditis (Figure 5). Nodules should be periodically observed, and if they are highly suspicious, FNA should be repeated. A recent study showed the result of repeat FNAs in 282 benign nodules.[38] Thirty-five of the initially benign nodules showed features suspicious of malignancy, and 7 (20%) of these suspicious nodules were proven to be malignant at surgery. The authors recommended a minimum of 3 benign FNA studies for complete follow-up; much more than 3 would certainly not seem practical. Indeed, if a clinician is so suspicious of malignancy that he pursues this route, then one could question why surgery was not recommended earlier. The value of repeated FNA in over 10,000 procedures showed that repeating thyroid FNAs on benign cytologically diagnosed nodules improved the accuracy in the diagnosis from 90% to 98%. The authors recommended repeating FNA in 1 year for patients with TNs diagnosed as benign.[39]

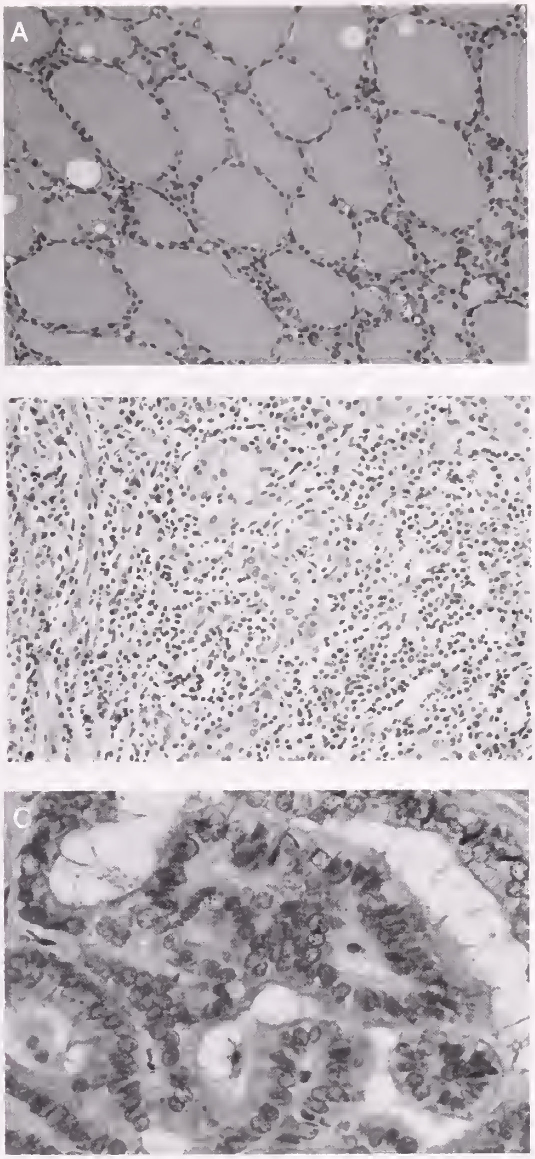


Fig 5. Pathological histological features of thyroid tissue: (A) normal thyroid (hematoxylin-eosin; original magnification, ×200), (B) Hashimoto's thyroiditis (note the marked lymphocytic infiltration; hematoxylin-eosin; original magnification, ×200), and (C) papillary carcinoma of the thyroid (note the intranuclear cytoplasmic pseudoinclusions and irregular orphan Annie nucleae; hematoxylin-eosin; original magnification, ×400) [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.].

Malignant This category encompasses carcinoma (papillary, medullary, and anaplastic), lymphoma, and metastases from another primary cancer. Cytological features of papillary thyroid cancer are straightforward, and generally the diagnosis is not an issue (Table 3 and Figure 5). The recommended procedure for these lesions is, of course, surgical excision.

Follicular Neoplasm Approximately 15% of TNs fall under this category, which covers both benign (adenoma) and malignant (carcinoma) follicular neoplasias. Around 20% of these cases are cancers.[20] Cytology cannot usually distinguish between follicular adenoma and follicular carcinoma, as the cells appear very similar. The pathological difference lies in the presence of vascular or capsular invasion, and this feature can be evaluated only by histological examination of the excised nodule. The recommended procedure for this category is uncertain. The clinician must decide whether to follow the patient for a possible increase in size or to recommend surgical excision. Such decisions are influenced by other suggestive features of potential malignancy discussed earlier.

Inconclusive or Indeterminate There are situations in which, despite the presence of adequate material, the cytological characteristics are not sufficient for diagnosis, as may occur in hyperplastic Hürthle cell nodules, or in which features do not permit distinguishing between adenomatous nodules and follicular neoplasms. The recommended procedure is to repeat the FNA or to recommend surgery if other clinical or sonographic suspicious features are present.

Suspicious for Carcinoma This category refers to cellular atypia, but the number and features of cells are insufficient for a definitive diagnosis of malignancy. It is recommended that surgery be advised in this situation.

Multiple Nodules

The management of patients with multiple TNs does not differ from the management of patients with solitary nodules. In the evaluation of multinodular goiters, it is advisable to biopsy nodules showing suspicious features on US as well as the larger nodules. The prevalence of malignancy in nodules from multinodular goiters is comparable to that in solitary nodules: 13% for both categories in 1 study.[8] Cancer was multifocal in 46% of such cases with multiple nodules, and 72% of cancers occurred in the largest (dominant) nodule. It can be concluded that in patients with 1 or more nodules, the likelihood of thyroid cancer per patient is independent of the number of nodules, whereas the likelihood per nodule decreases as the number of nodules increases. For cancer exclusion in a thyroid with multiple nodules larger than 1 cm, up to 4 nodules should be considered for FNA evaluation.[8]

Molecular Markers of Malignancy

Immunohistochemical markers for abnormal protein expression (eg, glucose transporters and galectin)

Table 3. *Cytological Categories of FNA Specimens with Suggested Management.*

Category	Diagnosis	Suggested Management	Frequency
Benign	Adenomatous nodule with abundant colloid Cyst, thyroiditis (chronic, acute, subacute, or fibrous), or multinodular goiter	Periodic observation including repeated FNA	70% (50%–90%)
Malignant	Papillary, medullary, or anaplastic carcinomas Lymphoma or rarely metastases	Surgical intervention: almost always total thyroidectomy	5% (1%–10%)
Follicular neoplasm	Follicular adenoma or follicular carcinoma Oxyphilic cell neoplasm	Surgical intervention: hemithyroidectomy or total thyroidectomy	10% (5%–20%)
Inconclusive or indeterminate	Cellular adenomatous nodule versus follicular neoplasm Cellular follicular lesion favoring neoplastic nodule Hyperplastic oxyphilic (Hürthle) cell nodule	Consideration of repeat FNA or surgical intervention: hemithyroidectomy or total thyroidectomy	10% (5%–15%)
Suspicious for carcinoma	Suspected papillary or medullary carcinoma Cellular atypia with insufficient cellularity	Consideration of surgical intervention	2% (0%–5%)
Inadequate	Insufficient material for diagnosis	Repetition of FNA	10% (5%–30%)*

Abbreviation: FNA, fine needle aspiration.
* This can be reduced by rebiopsy.

have been uniformly disappointing in comparison with direct cytology.[40] However, with the identification of V-raf murine sarcoma viral oncogene homolog B1 (Braf) and rearranged during transfection papillary thyroid carcinoma (Ret-PTC) mutations in up to 70% of papillary thyroid cancers,[41] the era of molecular diagnostics is upon us. Preliminary studies indicate that polymerase chain reaction analysis of FNA samples can contribute to the identification of malignancy.[42,43] However, to date, the sensitivity of such testing does not match the cytological assessment but may contribute to the identification of high-risk patients. In the near future, molecular markers will likely succeed the cytological approach, but they are not yet part of the routine TN evaluation.[5]

TREATMENT APPROACH

The treatment plan may depend on the patient's age and gender and the characteristics of the nodule (cytology, size, consistency, activity, and number). However, the diagnosis given by the cytologist is the most reliable indication for treatment. The clinical criteria should prevail only when cytological study is inconclusive.[34] However, it is fair to say that there remain significant differences between subspecialists in treatment preferences and that this reflects the lack of evidence-based medicine in this area.[44–48,50]

Asymptomatic benign euthyroid TNs need only a watchful follow-up with annual or biannual US and TSH testing. Nevertheless, on some occasions benign TNs do require therapeutic intervention, especially when they grow large and low and when they cause obstructive symptoms.

Any type of malignancy (or when malignancy cannot be ruled out) indicates the need for surgical resection. The extent of surgery depends, in some centers, on the type of tumor. Simple hemithyroidectomy may be recommended as the initial step for a follicular neoplasm or as the definitive treatment for a small, isolated papillary carcinoma if no other risk factors are present.[5] All other thyroid malignancies deserve total thyroidectomy with complementary local node resection. If the histology of the follicular neoplasm confirms the presence of carcinoma, then a total thyroidectomy should be completed.[5] The characteristics of the commonest treatments for TNs are summarized in Table 4.

Surgery

Indications

There are 4 major indications for thyroid surgery in patients with TNs:

1. Suspected malignancy.
2. Large goiters that generate local compressive symptoms.
3. Toxic nodules in young patients.
4. Recurrent cysts after ineffective therapeutic aspirations of the fluid because these lesions may have malignant cells in up to 10% of cases.[16]

Precautions

In addition to the routine potential complications of thyroid surgery, including hypoparathyroidism (temporary or permanent) and damage to the recurrent laryngeal nerve (temporary or permanent), long-term compression of the trachea may lead to atrophy and tracheal collapse after surgery.

Efficacy

Recurrence is common (15%–40%) in inadequately removed multinodular goiters.[51] A second intervention increases the chance of complications. Hence, we recommend a total thyroidectomy as the initial procedure in patients with bilateral nodules. Presurgical thyroxine administration has been suggested to reduce the gland size and consequently the surgical risk, but there is no evidence to support this approach. Also, postsurgical thyroxine is often prescribed with the idea of suppressing any future nodule formation. Although this is logical, the evidence for this maneuver is also absent.

Advantages

In addition to immediate relief of symptoms and concerns, surgery provides a gland for pathological evaluation.

Disadvantages

This is an expensive procedure that usually requires at least overnight hospitalization. The associated morbidity and efficacy are heavily dependent on the surgeon's experience. The main complications are lesions of the recurrent nerve (1%) and in total surgery the problem of permanent hypoparathyroidism (1%). The development of postsurgical hypothyroidism is dependent on the amount and quality of the remaining tissue (only ~1% with hemithyroidectomies).

Radioiodine ^{131}I

Indications

The main indications for radioiodine ablation are as follows: (1) hyperfunctioning multinodular goiters

Table 4. *Characteristics of the Commonest Treatments for a Thyroid Nodule.*

	Surgery	Radioiodine	Thyroxine
Main indications	Suspected malignancy Large goiter Toxic nodules Recurrent cyst	Hyperfunctioning nodule or multinodular goiter (especially in the elderly) Graves' disease	When no other treatment option is possible in a euthyroid patient
Contraindications	Surgical contraindications	Pregnancy and lactation	Low plasma TSH level
Efficacy	Excellent [recurrences are common with incomplete thyroidectomy (15%–40%)]	50% goiter reduction after 1 year	Poor Size reduction of 15%–40% but recurrence if treatment is stopped
Advantages	Immediate relief of compressive symptoms Availability of definitive histologic diagnosis	Medium-term relief of symptoms Outpatient procedure with low cost Well-tolerated Repeat treatment feasible	Low cost Possible prevention of new nodule formation
Disadvantages	High cost: required hospitalization Surgery-associated risks: vocal cord paralysis (1%) and hypoparathyroidism (1%) Hypothyroidism Recurrences possible	Secondary effects: hypothyroidism (50%–90%), thyroiditis (1%–3%), transient enlargement, and Graves' disease (1%–5%) Less effective in large goiters	Induction of osteopenia and atrial fibrillation Lifelong treatment
Other considerations		Nodule malignancy must be ruled out before treatment. Pretreatment with rhTSH may improve the outcome.	

Abbreviations: rhTSH, recombinant human thyroid-stimulating hormone; TSH, thyroid-stimulating hormone.

(for which it may be the treatment of choice); (2) hyperfunctioning single nodules; and (3) the need to reduce the size of a normally functioning multinodular goiter that is causing cosmetic or obstructive difficulties for the patient,[51] especially when surgery is contraindicated. Generally, RAI is a preferred therapeutic option for elderly patients.

Precautions

Before RAI administration, any possibility of malignancy should be ruled out. Pregnancy is of course a contraindication, but the fear of developing obstructive problems after the administration of the radioiodine is not justified by experience.[52] Radiation-induced transient thyroiditis, with a tender gland and increased thyroid hormone levels, may develop in the week following treatment of a toxic multinodular gland, but this is more often seen after the treatment of Graves' disease.

Dose

For RAI, 100 to 200 μCi of I^{131} therapeutic activity per gram of thyroid tissue is often recommended.[16,51] Generally, to ablate a gland, treatment with 20 to 30 mCi achieves >90% success.[51]

Efficacy

The success of radioiodine treatment depends on many factors other than the 24-hour radioiodine uptake. These include the administered dose and TN consistency because cystic and fibrotic areas will be more resistant to shrinkage. The decrease in nodule size after the procedure is directly related to the dose and inversely related to the initial nodule size. In small (<100 g) multinodular goiters, the typical reduction is about 50%, and the efficacy of the second dose is comparable to that of the first.[53] The shrinkage normally starts in the first weeks, but the maximum result may not be achieved for 1 to 2 years. Around 15% to 30% of patients require a second dose unless maximum ablation is the goal.

Advantages

This is a simple and cheap procedure that does not require hospitalization.

Disadvantages

If the physician is aiming for euthyroidism, he will be disappointed. The longer the patient is followed, the greater the likelihood of hypothyroidism developing. Using size-related doses causes mild hypothyroidism in 10% to 40% of cases, whereas overt hypothyroidism occurs in up to 50% of patients. Thyroid failure develops mostly in the first 2 years after treatment. Hypothyroidism is particularly frequent in those who have small goiters, anti–thyroid peroxidase antibodies, and/or a family history of thyroid autoimmunity.

Transient thyrotoxicosis, thyroid swelling, neck pain, and tenderness may be seen during the first month after RAI in 10% of cases, but postradiation inflammation and swelling that would require decompressive surgery are not seen. Long-term studies have shown that the associated risk of non–thyroid cancer development after RAI treatment is small.[54] However, around 1% to 5% of patients treated with lower doses of RAI may develop Graves' disease,[51] which is presumably precipitated by the release of thyroid autoantigens and other immunogenic effects of radioiodine on thyroid-autoreactive lymphocytes. Graves' disease with nodules is a known clinical entity of 2 common disorders developing together, and RAI may cause subclinical Graves' disease to worsen. This reaction is seen at 3 to 10 months after ^{131}I administration and is more frequent in anti–thyroid peroxidase–positive subjects.

Utility of Recombinant Human Thyroid-Stimulating Hormone (Thyrogen)

Pretreatment with recombinant human thyroid-stimulating hormone offers the prospect that radioiodine treatment of nontoxic nodular goiters can be simplified and improved, although more information about this strategy is needed.[51] However, this is particularly important in patients with multinodular goiters and low RAI uptakes who would be best treated with RAI. Some studies of nontoxic goiters have shown that pretreatment with 0.03 mg of recombinant human thyroid-stimulating hormone increases the RAI uptake to such an extent that it allows an RAI dose reduction of 60% and improves thyroid size reduction by 35% but with a 5-fold increase in the rate of early hypothyroidism.[53,55]

Other Options

Thyroxine

The rationale for this strategy is to suppress the TSH stimulatory effect on the TN by providing an excess of thyroxine in the circulation to suppress endogenous TSH levels. This strategy does not work well, and where it has been shown to work, any effect is lost after withdrawal of thyroxine.[56] Suppressing TSH in this way causes atrophy of the normal thyroid, and this may give the incorrect appearance of the nodule becoming smaller. Given the associated risks of high thyroid hormone levels (eg, arrhythmias and bone thinning), routine thyroxine use is not recommended because such side effects overcome the benefits.[56,57] There are also data suggesting that thyroxine can reduce multinodular goiter size by 15% to 40% in approximately 50% of subjects, but once again this is likely to be mostly due to a reduction in the normal thyroid tissue, and the same side effects are apparent.[51,58]

Ethanol

The injection, under US control, of ethanol into a TN may be a painful procedure. The size of the nodules may be reduced by 50% in 50% of patients, but usually this needs repeated sessions, sometimes up to 12.[59] The danger of local fibrosis and scarring of the recurrent laryngeal nerve has reduced enthusiasm for this procedure at the present time.

Laser Therapy

Ultrasonographically guided interstitial laser photocoagulation has been used as a nonsurgical alternative in the treatment of benign, solitary, solid (both cold and hot) TNs in patients who cannot (or will not) undergo surgery.[60,61] The technique has the advantage of not inducing hypothyroidism, but the efficacy is inferior to ^{131}I in normalizing TSH in patients with hot nodules. This procedure warrants further investigation.

The Follow-Up

The follow-up of benign euthyroid TNs depends on the clinical, US, and cytological characteristics described earlier. Usually, TNs require re-evaluation every 6 to 18 months by clinical assessment (history and physical examination), US, and TSH.[52] Nodules that have been cytologically labeled as inadequate, inconclusive, or indeterminate may need more careful monitoring and repeat biopsies.

Substantial growth detected by US (defined as an enlargement of more than 20% or more than 2 mm) does not necessarily indicate malignancy because benign lesions also grow larger, but a repeat FNA should be considered.[5,34] The same recommendation is valid if new clinical or US features of malignancy appear.

Thyroid function should also be regularly investigated (usually yearly) in those patients who have been treated with ablative options (surgery or ^{131}I). This is generally evaluated by a serum TSH determination. If hypothyroidism shows up, treatment should be started with thyroxine replacement therapy, which itself also requires annual follow-up. Postsurgical US is useful in monitoring the evolution of the remaining gland.

CONCLUSIONS

The adequate initial management of TNs requires a basic clinical and laboratory investigation including a US image and cytological assessment to rule out malignancy. Treatment should be individualized according to both the patient's and nodule's characteristics because there are several therapeutic options. Usually, a yearly evaluation is sufficient for long-term follow-up.

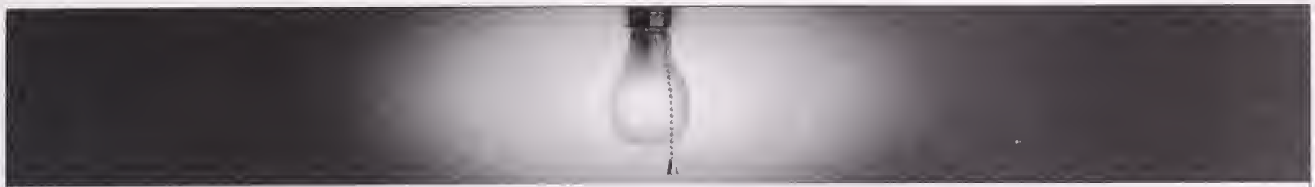
DISCLOSURES

Dr. Davies is a consultant of Kronus Corp, Boise, Idaho who distribute thyroid diagnostics.

REFERENCES

1. Gharib H, Papini E. Thyroid nodules: clinical importance, assessment, and treatment. *Endocrinol Metab Clin North Am* 2007; 36: 707–735.
2. The thyroid neck check. Available at: http://www.thyroidawareness.com. Accessed April 2008.
3. Davies T. Personalized thyroid medicine–the time is now. *Thyroid* 2007; 17: 1.
4. Mazzaferri EL. Management of a solitary thyroid nodule. *N Engl J Med* 1993; 328: 553–559.
5. Cooper DS, Doherty GM, Haugen BR, et al. Management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid* 2006; 16: 109–142.
6. Reiners C, Wegscheider K, Schicha H, et al. Prevalence of thyroid disorders in the working population of Germany: ultrasonography screening in 96,278 unselected employees. *Thyroid* 2004; 14: 926–932.
7. Burguera B, Gharib H. Thyroid incidentalomas. Prevalence, diagnosis, significance, and management. *Endocrinol Metab Clin North Am* 2000; 29: 187–203.
8. Frates MC, Benson CB, Doubilet PM, et al. Prevalence and distribution of carcinoma in patients with solitary and multiple thyroid nodules on sonography. *J Clin Endocrinol Metab* 2006; 91: 3411–3417.
9. Knudsen N, Laurberg P, Perrild H, et al. Risk factors for goiter and thyroid nodules. *Thyroid* 2002; 12: 879–888.
10. Kondo T, Ezzat S, Asa SL. Pathogenetic mechanisms in thyroid follicular-cell neoplasia. *Nat Rev Cancer* 2006; 6: 292–306.
11. Vassart G. Activating mutations of the TSH receptor. *Thyroid* 2004; 14: 86–87.
12. Kimura T, Van Keymeulen A, Golstein J, et al. Regulation of thyroid cell proliferation by TSH and other factors: a critical evaluation of in vitro models. *Endocr Rev* 2001; 22: 631–656.
13. Krohn K, Führer D, Bayer Y, et al. Molecular pathogenesis of euthyroid and toxic multinodular goiter. *Endocr Rev* 2005; 26: 504–524.
14. Alexander EK, Hurwitz S, Heering JP, et al. Natural history of benign solid and cystic thyroid nodules. *Ann Intern Med* 2003; 138: 315–318.
15. Erdogan MF, Gursoy A, Erdogan G. Natural course of benign thyroid nodules in a moderately iodine-deficient area. *Clin Endocrinol* 2006; 65: 767–771.
16. Hegedus L, Bonnema SJ, Bennedbaek FN. Management of simple nodular goiter: current status and future perspectives. *Endocr Rev* 2003; 24: 102–132.
17. Pacini F, Schlumberger M, Dralle H, et al. European consensus for the management of patients with differentiated thyroid carcinoma of the follicular epithelium. *Eur J Endocrinol* 2006; 154: 787–803.
18. Davies TF. Is consensus a good thing in the management of thyroid nodules? *Thyroid* 2006; 16: 205.
19. Burman KD. Micropapillary thyroid cancer: should we aspirate all nodules regardless of size? *J Clin Endocrinol Metab* 2006; 91: 2043–2046.
20. Pardal R, Clarke MF, Morrison SJ. Applying the principles of stem-cell biology to cancer. *Nat Rev Cancer* 2003; 3: 895–902.
21. Mitsutake N, Iwao A, Nagai K, et al. Characterization of side population in thyroid cancer cell lines: cancer stem-like cells are enriched partly but not exclusively. *Endocrinology* 2007; 148: 1797–1803.
22. Nam-Goong IS, Kim HY, Gong G, et al. Ultrasonography-guided fine-needle aspiration of thyroid incidentaloma: correlation with pathological findings. *Clin Endocrinol* 2004; 60: 21–28.
23. Herraiz M, Barbesino G, Faquin W, et al. Prevalence of thyroid cancer in familial adenomatous polyposis syndrome and the role of screening ultrasound examinations. *Clin Gastroenterol Hepatol* 2007; 5: 367–373.
24. Galofre JC. Management of subclinical hyperthyroidism. *Rev Med Univ Navarra* 2007; 51: 18–22.
25. Kraiem Z, Glaser B, Yigla M, et al. Toxic multinodular goiter: a variant of autoimmune hyperthyroidism. *J Clin Endocrinol Metab* 1987; 65: 659–664.
26. Boelaert K, Horacek J, Holder RL, et al. Serum thyrotropin concentration as a novel predictor of malignancy in thyroid nodules investigated by fine-needle aspiration. *J Clin Endocrinol Metab* 2006; 91: 4295–4301.
27. Vierhapper H, Niederle B, Bieglmayer C, et al. Early diagnosis and curative therapy of medullary thyroid carcinoma by routine measurement of serum calcitonin in patients with thyroid disorders. *Thyroid* 2005; 15: 1267–1272.
28. Elisei R, Bottici V, Luchetti F, et al. Impact of routine measurement of serum calcitonin on the diagnosis and outcome of medullary thyroid cancer: experience in 10,864 patients with nodular thyroid disorders. *J Clin Endocrinol Metab* 2004; 89: 163–168.

29. Baker JR Jr, Fosso CK. Immunological aspects of cancers arising from thyroid follicular cells. *Endocr Rev* 1993; 14: 729–746.
30. Matsubayashi S, Kawai K, Matsumoto Y, et al. The correlation between papillary thyroid carcinoma and lymphocytic infiltration in the thyroid gland. *J Clin Endocrinol Metab* 1995; 80: 3421–3424.
31. Frates MC, Benson CB, Charboneau JW, et al. Management of thyroid nodules detected at US: Society of Radiologists in Ultrasound consensus conference statement. *Radiology* 2005; 237: 794–800.
32. Rajatanavin R, Safran M, Stoller WA, et al. Five patients with iodine-induced hyperthyroidism. *Am J Med* 1984; 77: 378–834.
33. Kim JM, Ryu JS, Kim TY, et al. 18F-fluorodeoxyglucose positron emission tomography does not predict malignancy in thyroid nodules cytologically diagnosed as follicular neoplasm. *J Clin Endocrinol Metab* 2007; 92: 1630–1634.
34. Castro MR, Gharib H. Continuing controversies in the management of thyroid nodules. *Ann Intern Med* 2005; 142: 926–931.
35. Oertel YC. Unsatisfactory (vs. nondiagnostic) thyroidal aspirates: a semantic issue? *Diagn Cytopathol* 2006; 34: 87–88.
36. Oertel YC. Cytopathology reports from fine needle aspirations of the thyroid gland: can they be improved? *Thyroid* 2007; 17: 33–35.
37. Kelman AS, Rathan A, Leibowitz J, et al. Thyroid cytology and the risk of malignancy in thyroid nodules: importance of nuclear atypia in indeterminate specimens. *Thyroid* 2001; 11: 271–277.
38. Illouz F, Rodien P, Saint-Andre JP, et al. Usefulness of repeated fine-needle cytology in the follow-up of non-operated thyroid nodules. *Eur J Endocrinol* 2007; 156: 303–308.
39. Oertel YC, Miyahara-Felipe L, Mendoza MG, Yu K. Value of repeated fine needle aspirations of the thyroid: an analysis of over ten thousand FNAs. *Thyroid* 2007; 17: 1061–1066.
40. Haugen BR, Woodmansee WW, McDermott MT. Towards improving the utility of fine-needle aspiration biopsy for the diagnosis of thyroid tumours. *Clin Endocrinol* 2002; 56: 281–290.
41. Xing M. BRAF mutation in thyroid cancer. *Endocr Relat Cancer* 2005; 12: 245–262.
42. Foukakis T, Gusnanto A, Au AY, et al. A PCR-based expression signature of malignancy in follicular thyroid tumors. *Endocr Relat Cancer* 2007; 14: 381–391.
43. Delys L, Detours V, Franc B, et al. Gene expression and the biological phenotype of papillary thyroid carcinomas. *Oncogene* 2007; 26: 7894–7903.
44. Bonnema SJ, Bennedbaek FN, Wiersinga WM, Hegedus L. Management of the nontoxic multinodular goitre: a European questionnaire study. *Clin Endocrinol* 2000; 53: 5–12.
45. Bennedbaek FN, Hegedus L. Management of the solitary thyroid nodule: results of a North American survey. *J Clin Endocrinol Metab* 2000; 85: 2493–2498.
46. Bennedbaek FN, Perrild H, Hegedus L. Diagnosis and treatment of the solitary thyroid nodule. Results of a European survey. *Clin Endocrinol* 1999; 50: 357–363.
47. Bhagat MC, Dhaliwal SS, Bonnema SJ, et al. Differences between endocrine surgeons and endocrinologists in the management of non-toxic multinodular goitre. *Br J Surg* 2003; 90: 1103–1112.
48. Diehl LA, Garcia V, Bonnema SJ, et al. Management of the nontoxic multinodular goiter in Latin America: comparison with North America and Europe, an electronic survey. *J Clin Endocrinol Metab* 2005; 90: 117–123.
50. Walsh JP, Ryan SA, Lisewski D, et al. Differences between endocrinologists and endocrine surgeons in management of the solitary thyroid nodule. *Clin Endocrinol* 2007; 66: 844–853.
51. Weetman AP. Radioiodine treatment for benign thyroid diseases. *Clin Endocrinol* 2007; 66: 757–764.
52. Hegedus L. The thyroid nodule. *N Engl J Med* 2004; 351: 1764–1771.
53. Albino CC, Mesa CO Jr, Olandoski M, et al. Recombinant human thyrotropin as adjuvant in the treatment of multinodular goiters with radioiodine. *J Clin Endocrinol Metab* 2005; 90: 2775–2780.
54. Read CH Jr, Tansey MJ, Menda Y. A 36-year retrospective analysis of the efficacy and safety of radioactive iodine in treating young Graves' patients. *J Clin Endocrinol Metab* 2004; 89: 4229–4233.
55. Nielsen VE, Bonnema SJ, Hegedus L. The effects of recombinant human thyrotropin, in normal subjects and patients with goitre. *Clin Endocrinol* 2004; 61: 655–663.
56. Sdano MT, Falciglia M, Welge JA, Steward DL. Efficacy of thyroid hormone suppression for benign thyroid nodules: meta-analysis of randomized trials. *Otolaryngol Head Neck Surg* 2005; 133: 391–396.
57. Koc M, Ersoz HO, Akpinar I, et al. Effect of low- and high-dose levothyroxine on thyroid nodule volume: a crossover placebo-controlled trial. *Clin Endocrinol* 2002; 57: 621–628.
58. Mandel SJ, Brent GA, Larsen PR. Levothyroxine therapy in patients with thyroid disease. *Ann Intern Med* 1993; 119: 492–502.
59. Del Prete S, Caraglia M, Russo D, et al. Percutaneous ethanol injection efficacy in the treatment of large symptomatic thyroid cystic nodules: ten-year follow-up of a large series. *Thyroid* 2002; 12: 815–821.
60. Dossing H, Bennedbaek FN, Karstrup S, Hegedus L. Benign solitary solid cold thyroid nodules: US-guided interstitial laser photocoagulation-initial experience. *Radiology* 2002; 225: 53–57.
61. Dossing H, Bennedbaek FN, Bonnema SJ, et al. Randomized prospective study comparing a single radioiodine dose and a single laser therapy session in autonomously functioning thyroid nodules. *Eur J Endocrinol* 2007; 157: 95–100.
62. Davies TF, Schwartz AE. Thyroid nodules: malignant or benign? *Acute Care Med* 1985; 36–45.



don't publish online in the dark...

let our open linking network
shed some light.

This journal is CrossRef-ed! As a scientist or scholar navigating digital content, you've come to expect dynamic linking from citation to original source. As an author, you expect ready dissemination of your work to others. CrossRef is a collaborative, cross-publisher reference linking network that allows users to click on a citation and be taken directly to the full-text location. An independent, non-profit organization that employs open standards, CrossRef was established in cooperation with leading research publishers and the International DOI Foundation. Since the service went live in 2000, CrossRef has linked millions of articles from thousands of journals. Find out more at www.crossref.org, and make sure all your electronically published work is CrossRef-enabled.

crossref

CROSSREF.ORG | [illegible]

200 WHEELER ROAD, 6TH FLOOR, BURLINGTON, MA 01803 • 781-359-2442

Fresh Content
Sliced Any Way You Like It

Wiley InterScience® OnlineBooks™

Content
Thousands of Wiley book titles across science, technology, medicine, business, and finance

Access
Single chapter or full title
One-time fee, subscription, or Pay-Per-View options

Capabilities
Search at the chapter level
Full text presented in PDF format
Citation download feature
COUNTER-compliant usage statistics

Satisfy your library's evolving research needs and budget with a customized blend of new content and licensing options at **Wiley InterScience® OnlineBooks™**. Pick and choose from an expanding roster of thousands of Wiley titles available online via **Wiley InterScience** to create a dynamic electronic book collection. Select from our new flexible licensing models – one-time fee, subscription, or Pay-Per-View. Add books as you need them from our distinguished portfolio of classic and cutting-edge titles. Whichever combination you select, enjoy unlimited concurrent access to content across the spectrum of science, medicine, technology, business, and finance.

WILEY
InterScience®
DISCOVER SOMETHING GREAT

WILEY
1807-2007 KNOWLEDGE FOR GENERATIONS

www.interscience.wiley.com

VOL. 75, NO. 4
JULY/AUGUST 2008

MOUNT SINAI JOURNAL OF MEDICINE

A Journal of Translational and Personalized Medicine

Issue Theme

Type 1 Diabetes Mellitus: New Understanding of Etiology, New Approaches to Management

Introduction by Ronald Tamler, MD, Division of Endocrinology, Diabetes, and Bone Disease, Mount Sinai School of Medicine

Complications of Type 1 Diabetes Mellitus: New Molecular Findings

Kidney and Pancreas Transplantation in Type 1 Diabetes Mellitus

Immunology and Genetics of Type 1 Diabetes Mellitus

Type 1 Diabetes Mellitus: Primary, Secondary, and Tertiary Prevention

Toward Cell-based Cure for Diabetes Mellitus: Advances in Production and Transplantation of Beta Cells

Insulin Pump Therapy: Better Shot at Diabetes Control

Special Clinical Features

Treatment of Carotid Artery Occlusions: Stents vs Endarterectomy
EBM Case Study: Minimally Symptomatic Inguinal Hernia
Hepatitis C: Current and Future Therapies

WILEY-BLACKWELL

Discover
Something
Great

Distinctive Content

Content
Journals • Backfile Collections
Reference Works & Encyclopedias
OnlineBooks™
Current Protocols Lab Manuals
Evidence-Based Resources

Access
Single article/chapter or full title
One-time fee, subscription,
or Pay-Per-View options
Unlimited concurrent usage

Capabilities
E-mail Alerts • EarlyView®
Citation download & tracking
Remote access • Mobile Edition®

Promote the process of scientific discovery with a different breed of research system: **Wiley InterScience**®. Long distinguished by its scope and extensive coverage of emerging disciplines, Wiley's authoritative content spans the spectrum of scientific, medical, technical, and business specialties. The dynamic **Wiley InterScience** platform adds value to your research by providing simple desktop access to this preeminent digital library of journals, reference works, books, laboratory manuals, evidence-based databases, and more. Researchers reap the combined benefits of top-tier resources and up-to-the-minute functionality. Experience the difference **Wiley InterScience** can make at your institution today.

WILEY
InterScience®
DISCOVER SOMETHING GREAT

WILEY
1807-2007 KNOWLEDGE FOR GENERATIONS

www.interscience.wiley.com

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

Editor-in-Chief

Penny A. Asbell

Deputy Editors

Dennis Charney
Maria New
C. Warren Olanow

Associate Editors

Stuart Aaronson
David H. Adams
Erwin Bottinger
Eric Hollander
Ravi Iyengar
Mary Klotman
Paul Klotman
Ihor R. Lemischka
Fred Lublin
Michael L. Marin
Thomas G. McGinn
Ira Nash
Rosamond Rhodes
Hugh Sampson
Mary Sano
Anthony Squire
Fred Suchy
Savio Woo

Editorial Board

William J. Bremner	University of Washington
Thomas M. Coffman	Duke University
Dwight Evans	University of Pennsylvania
Gabriel Haddad	University of California, San Diego
Heidi E. Hamm	Vanderbilt University
Matti Häyry	University of Manchester
Eve J. Higginbotham	Morehouse School of Medicine
Katherine A. High	University of Pennsylvania
Tony Hope	University of Oxford
John H. Krystal	Yale University
Raju Kucherlapati	Harvard Medical School
Robert J. Lefkowitz	Duke University
Robert Lisak	Wayne State University
George Lister	University of Texas
Philip A. Marsden	University of Toronto
Charles Nemeroff	Emory University
Scott Rauch	Harvard Medical School
Bernard Roizman	University of Chicago
Alan F. Schatzberg	Stanford University
Inder Verma	The Salk Institute
Jeffrey A. Whitsett	University of Cincinnati
Allan Wolkoff	Albert Einstein College of Medicine
Noam Zohar	Bar-Ilan University

Guest Editor

Ronald Tamler, MD
Assistant Professor of Medicine,
Endocrinology, Diabetes and Bone Disease
Mount Sinai School of Medicine

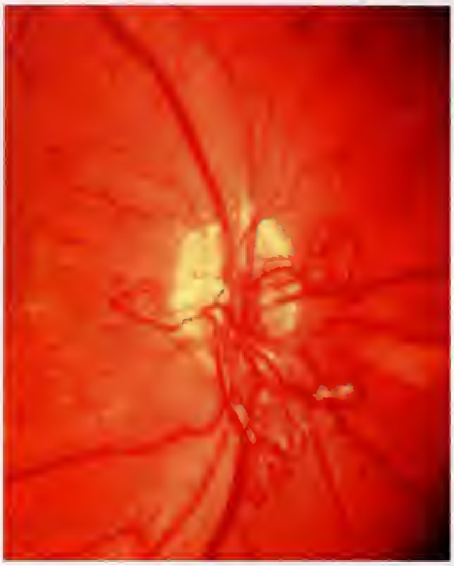
Managing Editor

Stephen Cohen
stephen.cohen@mssm.edu
212-241-6304

Publisher

John Wiley & Sons
Anna Wesley, Associate Editor
awesley@wiley.com

Proliferative Diabetic Retinopathy



This fundus photo shows a serious ocular complication of diabetes mellitus, consisting of the proliferative growth of abnormal new blood vessels at the optic disc. Appearing as a twisted collection of blood vessels, such neovascularization is dangerous because the abnormal vessels lead to vitreous hemorrhage, tractional retinal detachment, and loss of vision and even blindness. Reproduced by permission of Royal College of Opthalmologists, London, England.

Themes for future issues of *MSJM*, each shown with Guest Editor(s)

- Hospital medicine (Andrew Dunn)
- Advanced lung disease (Maria Padilla)
- Personalized medicine (Erwin Bottinger)
- Zoonotic infections (Stephanie Factor and Mary Klotman)
- Alzheimer's disease (Nilos Robakis)
- Traumatic brain injury (Andrew Jagoda)
- Medical education (Lisa Coplit)
- Drug discovery (Ravi Iyengar)
- Obstetrics/gynecology and reproductive services (Rhoda Sperling)

©2008 **Mount Sinai School of Medicine.** All rights reserved. No part of this publication may be reproduced in any form or by any means, except as permitted under section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the publisher, or authorization through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, Phone: (978) 750-8400; Fax: (978) 750-4770. Such requests and other permission inquiries should be addressed to the Permissions Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; Phone: (201) 748-6011; Fax: (201) 748-6008; E-mail: permreq@wiley.com.

Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine (Print ISSN: 0027-2507; Online ISSN: 1931-7581 at Wiley InterScience, www.interscience.wiley.com) is published twelve times per year for the Mount Sinai School of Medicine by Wiley Subscription Services, a Wiley Company, 111 River Street, Hoboken, NJ 07030.

Send subscription inquiries to: Subscription Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-6645.

Offprint sales and inquiries should be addressed to the Reprint Billing Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-8776. All other inquiries should be addressed to the Customer Service Department at (201) 748-6645.

Individual subscription price: $95.00 in North America, $120.00 outside North America. Payment must be made in US dollars drawn on a US bank. Claims for undelivered copies will be accepted only after the following issue has been delivered. For more subscription prices and information, please visit www.mountsinaijournal.org.

Change of address: Please forward six weeks prior to move: present mailing label with change of address. Please enclose a copy of the mailing label or cite your subscriber reference number in order to expedite handling. Missing copies will be supplied when losses have been sustained in transit and where reserve stock permits. Send claims in care of: *Mount Sinai Journal of Medicine*, c/o John Wiley & Sons, Inc., Attn: Journals Admin Dept. UK, 111 River Street, Hoboken, NJ 07030.

Cancellations: Subscription cancellations will not be accepted after the first issue has been mailed. Periodicals postage paid at Hoboken, NJ and at additional mailing offices.

Postmaster: Send address changes to *Mount Sinai Journal of Medicine*, Subscription Distribution, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Printed in the United States of America.

Advertising inquiries should be addressed to: Advertising Sales Manager, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Phone: (201) 748-6921; Fax: (201) 748-6207. Advertising in this journal is published with a caveat emptor understanding. The Publisher, Editors, and members of the Editorial Board do not imply warranty of products, nor quality, validity, or approval of education material offered by such advertising.

Indexed by: Biological Abstracts® (Thomson ISI), Chemical Abstracts Service/SciFinder (ACS), Current Contents®/Clinical Medicine (Thomson ISI), EMBASE/Excerpta Medica (Elsevier), Index Medicus/MEDLINE/PubMed (NLM), Science Citation Index Expanded™ (Thomson ISI), SCOPUS (Elsevier), Web of Science® (Thomson ISI).

This journal is printed on acid-free paper.

Disclaimer: All articles published, including but not limited to original research, reviews, clinical notes, clinical pearls, prismatic cases, clinician's corner, editorials, reports, and letters, represent the opinions and views of the authors and do not reflect any official policy or medical opinion of the Mount Sinai School of Medicine, the institutions with which the authors are affiliated, or the Publisher unless this is clearly specified. Articles published herein should not be relied upon as recommending or promoting a specific method, diagnosis, or treatment by physicians for any particular patient or disorder. Although the Editors and Publisher believe that drug selections and dosages and the specifications and usage of equipment and devices as set forth herein are in accord with current recommendations and practice at the time of publication, they accept no legal responsibility for any errors or omissions, and make no warranty, express or implied, with respect to material contained herein.

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

CONTENTS
VOL 75, NO 4

Theme Introduction

313 Type 1 Diabetes Mellitus: New Knowledge of Etiology, New Approaches to Management
Ronald Tamler

Theme Articles

314 Immunology and Genetics of Type 1 Diabetes
Michael P. Morran, Gilbert S. Omenn, Massimo Pietropaolo

328 Complications of Type 1 Diabetes: New Molecular Findings
Alin Stirban, Diethelm Tschoepe

352 Past, Present and Future of Insulin Pump Therapy: A Better Shot at Diabetes Control
Jennifer Sherr, Willielm V. Tamborlane

362 Toward Cell-based Cure for Diabetes: Advances in Production and Transplant of Beta Cells
Kathryn C. Claiborn and Doris A. Stoffers

372 Kidney and Pancreas Transplantation in Type 1 Diabetes Mellitus
Susan M. Lerner

385 Type 1 Diabetes Mellitus: Primary, Secondary, and Tertiary Prevention
Jennifer Bollyky, Srinath Sanda, Carla J. Greenbaum

Feature Departments

398 Controversy
Contemporary Issues in Management of Carotid Occlusive Disease
Gregg S. Landis, Peter L. Faries

401 Evidence-based Medicine (Editor: Thomas McGinn)
EBM on Call: Management of Minimally Symptomatic Inguinal Hernia
Deborah Korenstein, Thomas McGinn

405 New Drugs and the FDA (Editor: Jeffrey Newcorn)
Hepatitis C: Current and Future therapies
Poonam Mishra, Donald M. Jensen

THEME INTRODUCTION

Type 1 Diabetes Mellitus: New Knowledge of Etiology, New Approaches to Management

Ronald Tamler, MD, PhD

Division of Endocrinology, Diabetes, and Bone Disease, Mount Sinai School of Medicine, New York, NY

Type 1 diabetes mellitus (T1DM) is usually viewed as the "little brother" of epidemic type 2 diabetes; a mere tenth of patients afflicted with diabetes have type 1. However, much of what we know about diabetes and its complications in a more general sense stems from knowledge that we have gleaned from T1DM, starting with the strong impact of the Diabetes Control and Complications Trial and its follow-up studies.

The care of patients with T1DM epitomizes the principle of translational and personalized medicine to which this journal is dedicated. Epidemiologic and clinical data help bench researchers find out more about the immunology and pathophysiology of the disease and its comorbidities.

Meanwhile, the enormous strides made in the identification of possible targets for preventing and treating the disease and its complications have led to significant discoveries. These innovations are now being brought back to the clinical arena with revolutionary approaches in transplant medicine, immunology, and device technology.

Finally, as every clinician caring for patients with T1DM knows, medicine cannot get any more personalized than diabetes care when these novel technologies are finally applied to the patient's benefit.

The circle of learning and applied knowledge is followed in this issue. Morran, Omenn, and Pietropaolo introduce us to the immunology and genetics of T1DM, which consequently influences islet cell biology as summarized by Claiborn and Stoffers. Bollyky, Sanda, and Greenbaum then address different levels of prevention of T1DM, whereas Stirban, Rosen, and Tschoepe discuss complications and the mechanisms by which they arise. Finally, the reader learns about very concrete and current methods of treatment. Lerner reviews the great strides that have been made in pancreas transplantation, one of the great strengths of Mount Sinai Hospital. Sherr and Tamborlane point the way to the closed loop system of the "artificial pancreas" in their discussion of insulin pump devices and continuous glucose monitoring.

I would like to thank the editors of the *Mount Sinai Journal of Medicine* for recognizing the rapid development of knowledge in this field by dedicating a full issue to T1DM, and I want to thank our contributors from near and far for sharing their expertise with us. The great strides that are reported in these pages give rise to the hope that someday–hopefully soon–we will be able to report a cure.

Address Correspondence to:

Ronald Tamler
Division of Endocrinology,
Diabetes, and Bone Disease
Mount Sinai School of Medicine
New York, NY
Email: ronald.tamler@
mssm.edu

DISCLOSURES

Potential conflict of interest: Nothing to report.

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20059

© 2008 Mount Sinai School of Medicine

Immunology and Genetics of Type 1 Diabetes

Michael P. Morran,[1] Gilbert S. Omenn[2] and Massimo Pietropaolo[1]

[1] Laboratory of Immunogenetics, Brehm Center for Type 1 Diabetes Research and Analysis, Division of Metabolism, Endocrinology, and Diabetes, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI
[2] Center for Computational Medicine and Biology, Departments of Internal Medicine and Human Genetics, University of Michigan, Ann Arbor, MI

ABSTRACT

Type 1 diabetes is one of the most well-characterized autoimmune diseases. Type 1 diabetes compromises an individual's insulin production through the autoimmune destruction of pancreatic β-cells. Although much is understood about the mechanisms of this disease, multiple potential contributing factors are thought to play distinct parts in triggering type 1 diabetes. The immunological diagnosis of type 1 diabetes relies primarily on the detection of autoantibodies against islet antigens in the serum of type 1 diabetes mellitus patients. Genetic analyses of type 1 diabetes have linked human leukocyte antigen, specifically class II alleles, to susceptibility to disease onset. Environmental catalysts include various possible factors, such as viral infections, although the evidence linking infections with type 1 diabetes remains inconclusive. Imbalances within the immune system's system of checks and balances may promote immune activation, while undermining immune regulation. A lack of proper regulation and overactive pathogenic responses provide a framework for the development of autoimmune abnormalities. Type 1 diabetes is a predictable and potentially treatable disease that still requires much research to fully understand and pinpoint the exact triggering events leading to autoimmune activation. *In silico* research can aid the comprehension of the etiology of complex disease pathways, including Type I diabetes, in order to and help predict the outcome of therapeutic strategies aimed at preserving β-cell function. ***Mt Sinai J Med 75:314–327, 2008.*** © *2008 Mount Sinai School of Medicine*

Key Words: adaptive immunity, autoimmunity, genetics in type 1 diabetes, innate immunity, type 1 diabetes.

Address Correspondence to:

Massimo Pietropaolo
Laboratory of Immunogenetics
Brehm Center for Type 1
Diabetes Research and Analysis
Ann Arbor, MI
Email: maxtp@umich.edu

IMMUNE SYSTEM REVIEW

The immune system is the body's natural defense system against invading pathogens. It protects the body from infection and works to communicate an individual's well-being through a complex network of interconnected cells and cytokines. This system has the power to initiate a wide gamut of cellular responses with the ability to directly attack an invading organism or signal cells to begin the healing process. Although this system is an associated host defense, an uncontrolled immune system has the potential to trigger negative complications in the host. Therefore, well-controlled regulation of the immune system is necessary in order to prevent autoimmune responses from occurring.

In order to protect the body against foreign pathogens, the immune system has developed throughout evolution to recognize the difference between the self and nonself. The ability to become self-tolerant toward the body's own proteins and antigens is critical to maintaining a properly functioning immune system. An immune system that loses tolerance to the self loses the ability to differentiate between friend and foe in immunological battles. Loss of tolerance leads the immune system toward autoimmune responses, in which the body attacks itself, causing substantial damage to the self, even inflicting irreversible damage.

The immune system is composed of 2 unique branches, each with its own responsibilities. The

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20052

© 2008 Mount Sinai School of Medicine

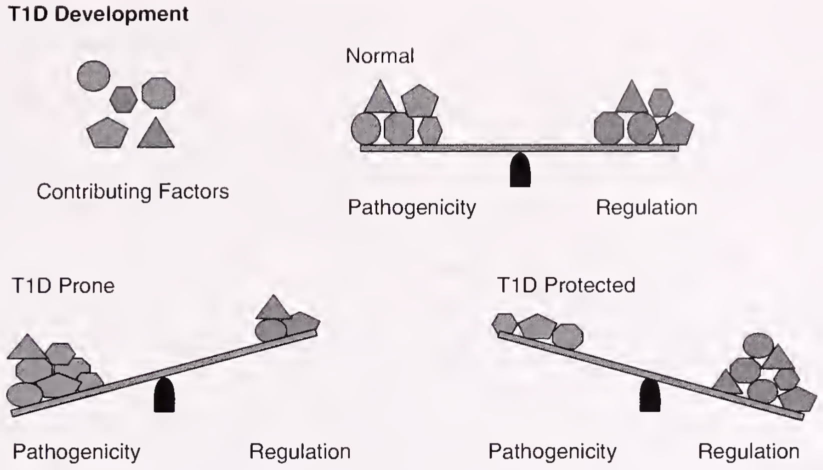


Fig 1. T1D development. A discrete balancing act is performed by the immune system in order to attack foreign pathogens while protecting against the promotion of autoimmunity. Multiple contributing factors influence the development of T1D, including genetics, the environment, and cellular signals. In a normal individual the immune system is balanced, possessing the ability to both promote and suppress cellular responses to pathogens. When an immune system is unbalanced and favors inflammation or excessive cellular responses, the system is unbalanced and favors excesive regulation of cellular responses, the system is protected against the development of T1D and autoimmunity. **Abbreviations:** T1D, type 1 diabetes. Adapted from *Annual Review of Immunology*.[102] [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.].

innate immune system is the body's first line of defense against invading pathogens. This system recognizes common structural components of pathogens and elicits immune responses to signal the presence of pathogens and infection.[1] The adaptive immune system is the body's secondary line of defense and specifically targets identified pathogens. This system is antigen-specific and generates immunological memory within the host, which allows for more efficient pathogen clearance upon repeat exposure to the same pathogen.[1] Although these 2 systems are termed different branches of the immune system, they must work together as one unified system to protect the body. The 2 branches of the immune system rely on each other to help properly perform their jobs. If either branch fails to perform its job, the other branch suffers. Regulatory malfunctions in either system can ultimately lead to the generation of unwarranted or unregulated autoimmune responses. Underlying an immune response lies a delicate network of cell types and cytokines that communicate to begin, generate, and end an immune response. If an immune system becomes unbalanced, this delicate network, responsible for preventing autoimmunity, promotes it (Figure 1).

The immune system protects the host, but it also possesses the power to harm the host as well. Numerous autoimmune diseases have been characterized over time. This article provides an overview of how alterations in the immune system, due to genetics, cellular malfunctions, or cell signaling issues, lead to the development and pathogenesis of autoimmune diseases, specifically type 1 diabetes (T1D). Understanding mechanistically how T1D and autoimmune responses are triggered is essential in order to develop strategies to combat these diseases.

TYPE 1 DIABETES OVERVIEW

Diabetes mellitus is a group of diseases characterized by the body's inability to accurately maintain normal blood glucose levels, leading to multiple detrimental effects. Insulin is an important hormone in glucose metabolism. When released, it signals cells to take up glucose. If the body is unable to produce insulin, blood glucose levels remain elevated, and this is termed *hyperglycemia*. T1D is an autoimmune disease in which the immune system targets and destroys the insulin-producing β-cells found in the islets of Langerhans in the pancreas. Without insulin, individuals develop the clinical syndrome of T1D.

T1D is characterized by autoantibody production and progressive infiltration of immune cells into the islets of the pancreas, followed by the destruction of the islet cells.[2] During the onset of T1D, cells from both the innate and adaptive immune systems infiltrate islet lesions to produce insulitis. Studies using human and murine models of diabetes have demonstrated that the autoimmune destructive process in T1D occurs in a cell-mediated organ-specific manner and requires both CD+4 and CD+8 T cells as well as macrophages.[3–13] These cells accumulate in an islet lesion but are nondestructive.[14–16] An unknown triggering event occurs, which promotes the autoimmune destruction of the β-cells.[14–16] Clinical manifestations of diabetes occur after 90% of an individual's β-cell mass is destroyed.[14]

The exact trigger for the onset of T1D is still unknown, as discussed later, although the mechanism by which the insulin-producing β-cells are destroyed is well understood. The destructive process is T cell–mediated. Once the islets have become infiltrated and highly populated with T cells and macrophages, a subsequent triggering event occurs to activate these cells.[14–16] T cells, B cells, and macrophages communicate via antigen presentation and can, in turn, activate each other via cytokines and direct cell communication through surface receptors.[1] With such a high population of immune cells centered in one distinct area, activation signals can travel fast, initiating a destructive cascade easily.

In general, macrophages play the role of antigen-presenting cells (APCs), and so they can directly pick up and present β-cell antigens. This leads to the activation of β-cell–cytotoxic CD+8 T cells and the generation of autoreactive CD+4 effector T cells.[15] Cytotoxic T cells can directly kill β-cells, whereas effector T cells can initiate the activation of B cells, thus prompting autoantibody production.[15] Once the immune system is triggered or activated, it begins to process and present self antigens to T cells, generating the autoimmune process that leads progressively to the clinical manifestations of T1D.

T cells possess the ability to directly destroy β-cells in a cytotoxic manner but also hold the power to directly influence the induction of β-cell destruction through the release of cytotoxic molecules, including cytokines, granzyme B, and perforin.[16,17] T cells can also signal β-cell death through the Fas pathway.[16,17] β-Cell apoptosis is promoted by activation of the caspase pathway through multiple mechanistic pathways, including Fas and Fas ligand interactions, nitric oxide and oxygen-derived free radicals, and membrane disruption due to perforin or granzyme B.[18] Cytokine production by activated T cells may be a key factor in β-cell death, influential cytokines released by T cells include: interleukin 1 (IL-1), interferon γ (IFNγ), and tumor necrosis factor α (TNFα), all of which up-regulate Fas and the production of Fas ligand, nitric oxide, and free-radicals.[18,19] Other notable cytokines are IL-2, IL-12, IL-17, and IL-18, all of which seem to promote an inflammatory state biased to T helper 1 type responses.[18,19]

AUTOANTIGENS IN TYPE 1 DIABETES

During the progression of T1D, multiple autoimmune processes occur. Human and murine studies of T1D have illustrated that, as the severity of disease increases, so does the number of autoantigens and autoantibodies.[20] Over the course of development, β-cells go through multiple divisions and changes in cell volume and increase in cell mass.[21] During this time, β-cells grow, develop, and eventually die; these dead cells may be engulfed and processed by APCs. In an insulitis lesion, these APCs presenting processed β-cell components have the potential to trigger the generation of β-cell autoreactive T cells.[14–16,21]

Most well-known autoantigens are associated with β-cell components, including insulin, glutamic acid decarboxylase (GAD), and islet-cell antigen-2 (IA-2).[20] Insulin is the first antigenic target detectable during the early progression of diabetes,[22] although most autoantibodies are targeted against the β-cells themselves and other β-cell–secreted proteins.[20] Recently, ZnT-8, a pancreatic β-cell–specific zinc transporter, has been identified as a candidate autoantigen.[23] Although research is limited on this molecule, strong evidence already supports the association of ZnT-8 autoantibodies with T1D.[23] During the progression of T1D, a process of autoantigen epitope spreading occurs.[24] Epitope spreading provides an explanation of how the immune system is capable of recognizing increasing numbers of autoantigens in correlation with increased T1D disease severity.[24] Epitope spreading begins with the immune system recognizing and mounting an immune response against a single antigen, which is recognized via a single epitope. Over time, new antigens can be recognized, and previously recognized antigens can be differentially processed by APCs to generate multiple epitopes for a single antigen.

One can think of this process as similar to a tree growing toward autoimmunity. In Figure 2, the tree stem symbolizes an immune system at birth that lacks autoimmunity. As this tree grows

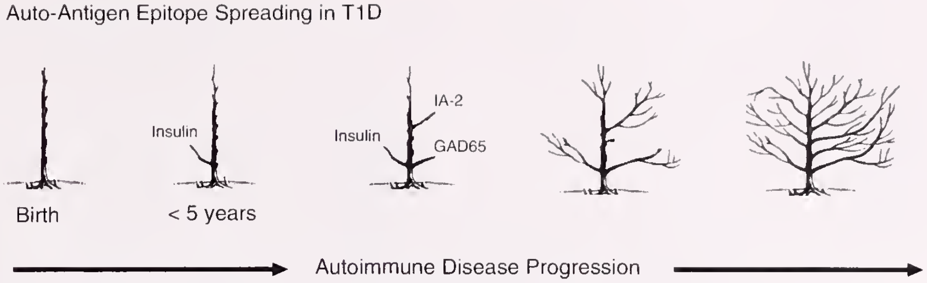

Auto-Antigen Epitope Spreading in T1D
Insulin
IA-2
Insulin
GAD65
Birth
< 5 years
Autoimmune Disease Progression

Fig 2. Autoantigen epitope spreading in T1D. As the severity of symptoms associated with T1D increases over time, so does the number of autoantigens recognized by the immune system. Epitope spreading begins once the immune system is triggered within the pancreas, leading to the processing and presentation of self antigens. As β-cell destruction takes place, multiple self antigens become targets of the immune system. During this process, insulin is the first antigenic target,[22] and it is followed by other β-cell–associated components, such as GAD65 and IA–2.[20] Over time, autoantigens are processed differently, creating various recognition epitopes for a given antigen. In Figure 1, the tree symbolizes an immune system at birth that lacks autoimmunity. As the tree grows toward autoimmune T1D, its limbs represent targeted self antigens that develop. As T1D progresses, multiple limbs grow off the tree, each from a different antigen. These growing limbs next branch off, representing the unique epitopes recognized from differential processing of similar self peptides. As T1D develops, the tree grows toward autoimmunity by increasing both the number of limbs and the number of branches on a given limb, representing the process of epitope spreading observed in T1D disease development. **Abbreviations:** GAD65, glutamic acid decarboxylase 65; IA-2, islet-cell antigen-2; T1D, type 1 diabetes.

toward autoimmune T1D, its limbs represent developing targeted self antigens. Next, multiple limbs grow from the stem, each targeting different antigens. These growing limbs branch off, representing the unique epitopes recognized from differential processing of similar peptides. It has been shown that autoantibody production has been detected up to 5 years prior to the development of hyperglycemic events, and this indicates that autoantibody production precedes the clinical manifestation of T1D.[24] As T1D develops, the tree grows toward autoimmunity by increasing both the number of limbs (targeted antigens) and the number of branches on a given limb (recognized epitopes for a particular antigen). Although the progression and mechanisms behind the development of T1D are understood in general terms, many questions remain to be answered concerning how the autoimmune disease state becomes triggered toward autoreactivity.

We have provided evidence suggesting that a subset of cytoplasmic islet-cell antibodies (ICAs) is related to a more rapid progression to insulin-requiring diabetes in glutamic acid decarboxylase 65 (GAD65) and IA-2 antibody–positive relatives of proband patients versus relatives with GAD65 and IA-2 antibodies without ICAs (Figure 3).[25] The precise nature of the antigen(s) detected by the indirect immunofluorescence ICA assay remains enigmatic, but its utility to predict future development of diabetes mellitus in individuals with other circulating ICAs has recently been confirmed.

WHAT ARE THE GENES ASSOCIATED WITH TYPE 1 DIABETES?

Although it has been suggested that multiple genes play a role in disease susceptibility, there is strong evidence for only 2 chromosomal regions that are associated with T1D: the human leukocyte antigen (HLA) region on chromosome 6p21 [insulin-dependent diabetes mellitus 1 (*IDDM1*)] and the insulin gene region on chromosome 11p15 [insulin-dependent diabetes mellitus 2 (*IDDM2*)]. The contributions of these 2 loci to familial inheritance are approximately 42% for *IDDM1* and 10% for *IDDM2*. As a result of genome-wide searches, many other putative loci have been proposed to be related to T1D. These loci along with some potential candidate genes are listed in Tables 1 and 2. The fact that in humans the highest risk-conferring locus linked to the disease is the HLA cluster and, in particular, HLA genes encoding specific class II alleles strongly indicates an important role of the immune cells in both the development and activation of the autoimmune response leading to disease onset.[26]

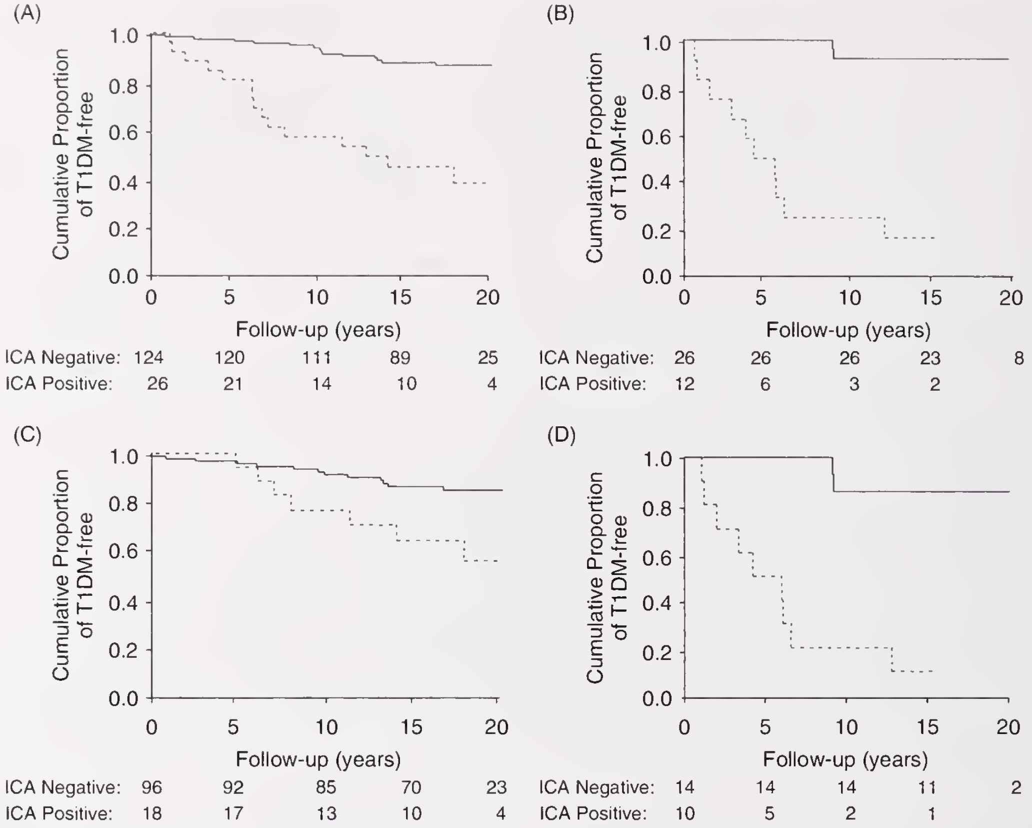


Fig 3. The rate of progression to T1DM development in relatives carrying GAD65 AA, IA-2 AA, or a combination of both AA in the (–) absence or (- - -) presence of ICA. (A) Progression to insulin-requiring diabetes for relatives with GAD65 AA with respect to ICA positivity. (B) Progression to insulin-requiring diabetes for relatives with IA-2 AA with respect to ICA positivity. (C) Progression to insulin-requiring diabetes for relatives with either GAD65 or IA-2 AA with respect to ICA positivity (log rank: $P = 0.01$). (D) Remarkably, the cumulative risk of developing an insulin requirement was 80% at 6.7 years and 90% at 12.9 years of follow-up in ICA-positive relatives; this is significantly higher than the cumulative risk of diabetes development in relatives who were positive for GAD65 and IA-2 AA without ICA (log rank: $P < 0.00001$). **Abbreviations:** AA, auto-antibodies GAD65, glutamic acid decarboxylase 65; IA-2, islet-cell antigen-2; ICA, islet-cell antibody; T1DM, type 1 diabetes mellitus. Reprinted with permission from *Pediatric Diabetes*.[25] Copyright 2005, International Society for Pediatric and Adolescent Diabetes.

Interestingly, the same HLA locus seems to have the corresponding susceptibility influence in the primary mouse model of T1D, the nonobese diabetic (NOD) mouse (Table 3).[27] The immune-mediated processes of β-cell destruction are mainly T cell–dependent and chronic in both mouse and rat models of T1D, and this makes it more likely to be the same in humans.[27] Comparative mapping of human (IDDM) and NOD mouse insulin-dependent diabetes (*Idd*) genes are shown in Table 3.

Genetic factors have long been thought to be linked to the development of T1D. Although it has been hypothesized that in monozygotic twins a discordance rate greater than 50% could be explained by environmental factors of disease development, studies using both monozygotic and dizygotic twins have suggested that environmental factors have few causative roles in the development of islet autoimmunity, whereas genetic similarities in dizygotic twins seem more important in determining susceptibility to diabetes.[28] T1D is a polygenic disease[29] in which there presumably exist a small number of genes with large effects, HLA being the main example, and a large number of genes with small effects overall.[30]

THE HUMAN LEUKOCYTE ANTIGEN COMPLEX

The short arm of human chromosome 6 (6p21) accommodates a ~ 3.5–megabase genetic segment containing a group of immune response genes

Table 1. *Effect of Human Leukocyte Antigen Alleles on Type 1 Diabetes Susceptibility.*

DQ Alleles	Effect	Associated DR
B1* 0302, A1* 0301	Susceptible	DR4
B1* 0201, A1* 0501	Susceptible	DR3
B1* 0501, A1* 0101	Susceptible	DR1
B1* 0201, A1* 0301	Susceptible (African Americans)	DR7
B1* 0502, A1* 0102	Susceptible (Sardinia)	DR2 (DR16)
B1* 0303, A1* 0301	Susceptible (Japanese)	DR4
B1* 0303, A1* 0301	Susceptible (Japanese)	DR9
B1* 0602, A1* 0102	Protective	DR2 (DR15)
B1* 0301, A1* 0501	Protective	DR5
B1* 0201, A1* 0201	Neutral	DR7
B1* 0303, A1* 0301	Neutral	DR4
B1* 0301, A1* 0301	Neutral	DR4

Table 2. *Genetic Risk Estimates for Human Leukocyte Antigen Class II in T1D.*

High-Risk Genotype	Risk in an Individual with This Genotype
DQB1* 0302 (DQ3.2)	1 in 60
DQ3.2/DQ2 (DR3)	1 in 25
DQB1* 0302 and a family history of IDDM	1 in 10
DQ3.2/DQ2 (DR3) and a family history of T1D	1 in 4

NOTE: This table was adapted from *Annual Review of Medicine*.[26]
Abbreviations: IDDM, insulin-dependent diabetes mellitus; T1D, type 1 diabetes.

termed the *major histocompatibility complex* (MHC). The principal genes located within the MHC code for HLAs, 2 molecular classes of cell surface glycoproteins differing in structure, function, and tissue distribution.

The class I HLA molecule exists as a heterodimer, consisting of a polymorphic 44-kDa MHC-encoded α or heavy chain in noncovalent association with β_2-microglobulin, a 12-kDa protein encoded by a nonpolymorphic gene on chromosome 15. The class I molecule is anchored in the cell membrane only by the heavy chain. This chain contains 338 amino acids and, beginning from the amino terminus, is functionally divided into 3 regions: an extracellular hydrophilic region, a transmembrane hydrophobic region, and an intracytoplasmic hydrophilic region. The extracellular region is further subdivided into 3 domains, designated α_1, α_2, and α_3, each of which has approximately 90 amino acid residues. The α_1 and α_2 domains compose the peptide- or antigen-binding region of the molecule.

Class II HLA molecules consist of 2 glycoprotein chains, an α chain of approximately 34 kDa and a βchain of approximately 29 kDa, both encoded within the MHC. As with the class I heavy chain, each class II chain can be divided into 3 regions (extracellular, transmembrane, and intracytoplasmic), but in contrast, both class II chains span the cell membrane. Each extracellular region of the class II α and βchains has been further divided into 2 domains of approximately 90 amino acid residues each, termed α_1 and α_2 and β_1 and β_2, respectively. The α_1 and β_1 domains form the peptide-binding region of class II HLA molecules. Also of note, the class II α_2 and β_2 domains, class I α_3 domain, and β_2-microglobulin all show homology to the constant region of immunoglobulins and are therefore classified as members of the immunoglobulin superfamily.

The genes that encode class I MHC are located at the HLA-A, HLA-B, and HLA-C loci, whereas class II molecules are encoded by the DR, DQ, and DP genes. Other genes in this cluster include transporters associated with antigen processing[31,32] and low-molecular-weight proteins, both of which are involved in antigen processing.[33] A third region of the MHC, termed class III, encodes several molecules having a variety of functions, such as complement

Table 3. *Comparative Mapping of Human (IDDM)and Murine (Idd) Genes.*

Human Locus (Chromosome)	Marker/ Candidate	Potential NOD "Idd" Homolog (Chromosome)	Marker
IDDM1 (6p21)	HLA	"*Idd1*" (17)	H2g7
IDDM2 (11p15.5)	INS/VNTR	? NOD.DR-2 (chromosome 7 region from C57L)	? Ins2
IDDM3 (15q26)	D15S107	"*Idd2*" (9)	Cyp19
IDDM4 (11q13)	FGF3	None yet identified (distal 7)	
IDDM5 (6q25)	ESR	None yet identified (proximal 10)	
IDDM6 (18q)	D18S64	? "*Idd5*" (1)	Bcl2
IDDM7 (2q31-33)	D2S326	? "*Idd5*" (1)	Il1r/Stat1
IDDM8 (6q25-27)	D6S264	None yet identified (proximal 10 or 17)	
IDDM9 (3q21-q25)	D3S1303	None yet identified (middle 6)	
IDDM10 (10p11.2-q11.2)	GAD2	None yet identified (proximal 2)	
IDDM11 (14q24.3-q31)	D14S67	None yet identified (middle 12)	
IDDM12 (2q31-33)		? "*Idd5.1*" (1)	Ctla4
IDDM13 (2q34)	IGFBP-2,5	? "*Idd5.2*"(1)	Slc11a1 (Nramp)
GCK (7p)	GCK	None yet identified (proximal 11)	
IDDM15 (6q21)	D6S283	? "*Idd14*" (13)	D13Mit61
IDDM16? (1p36.1-p35)	NHE1	? "*Idd11*" (4)	Slc9a1 (Nhe1)

NOTE: This table was adapted from *Molecular Pathology of Insulin Dependent Diabetes Mellitus.*[27] *Idd* nomenclature is placed in quotation marks because there are multiple Idd loci on mouse chromosome 1.
Abbreviations: GCK, glucokinase; HLA, human leukocyte antigen; Idd, insulin-dependent diabetes; IDDM, insulin-dependent diabetes mellitus; NOD, nonobese diabetic.

components (C4A, C4B, factor B, and C2), tumor necrosis factor α and tumor necrosis factor β, and the 21-hydroxylase genes (CYP21P and CYP21).

Polymerase chain reaction studies[34] have provided researchers with a rapid means of estimating T1D susceptibility in comparison with serological techniques. Polymerase chain reaction amplification of individuals' HLA alleles in a variety of racial and ethnic groups has revealed that the presence of a specific human DQβchain variant encoding a neutral amino acid (alanine, valine, or serine), rather than aspartic acid at position 57 (non-Asp-57), is strongly associated with T1D. In contrast, negatively charged aspartic acid at position 57 of the DQβchain (Asp-57) appears to confer resistance to T1D progression. This association is much stronger than the association between HLA-DR3 and HLA-DR4 and the presence of the disease (Tables 1 and 2).[35–37]

Susceptibility to T1D is mainly conferred by specific polymorphic regions within the MHC complex, such as HLA DR/DQ alleles.[35,36] The genotype associated with the highest risk for T1D is the DR3/4-DQ8 (DQ8 is DQA1*0301, DQB1*0302) heterozygous genotype. The HLA genotype DQB1*0602 confers dominant protection against T1D.

There is evidence that risk for islet autoimmunity increases in DR3/4-DQ8 siblings who share both HLA haplotypes with their diabetic proband sibling (63% by age 7 and 85% by age 15) in comparison with siblings who do not share both HLA haplotypes with their diabetic proband sibling.[38] These findings indicate that HLA genotyping at birth may identify individuals at high risk of developing the disease with no detectable signs of islet autoimmunity, who could then be enrolled in intervention trials aimed at preventing overt disease.

The mechanisms by which the class II genes can influence susceptibility to or protection from T1D are still the subject of discussion. Brown *et al.*[39,40] characterized the structure of the crystallized HLA class II molecule. One hypothesis is that effective antigen binding depends on the conformation of the antigen-binding site on the DQ dimer. The 2 critical residues, DQα Arg-52 and DQβAsp-57, are located at opposite ends of the α-helices that form the antigen-binding site of the DQ molecule. It has been postulated that a substitution of an amino acid residue at these positions of the DQ molecule leads to conformational changes of the antigen-binding site and consequently to a modification of the affinity of the class II molecule for the diabetogenic peptide(s).[37] As support for this hypothesis, it is known that Asp-57 is involved in hydrogen and salt bonding with both the peptide main chain and the DRα Arg-76 side chain. Theoretically, modifications in the DRα Arg-76 residue would also alter the antigen-binding site. It is noteworthy that studies of the regulatory regions of the genes encoding DQ α and DQ βchains have shown that the level of transcription of these genes may influence the amount of antigen binding. An increased level of production of a class II chain may increase the availability[41] for dimerization.[42] Studies by Demotz *et al.*[43] have suggested that relatively few class II heterodimers need to be present on the surface of an APC to efficiently crosslink the T cell receptor (TCR) and initiate a T cell response.

ENVIRONMENTAL FACTORS INFLUENCING TYPE 1 DIABETES

Although T1D has long been considered to be directly linked to genetic factors, researchers have been searching for environmental factors capable of triggering T1D progression. Epidemiologists have examined trends for environmental triggers through major geographical variations in T1D incidence, temporal trends in the incidence of T1D, and migration studies.[44] The responsibility of environmental factors for T1D is strongly supported by statistical evidence showing that the incidence of newly diagnosed T1D subjects with high-risk HLA genotypes has decreased over the last decades, whereas newly diagnosed T1D with low-risk or even protective HLA genotypes has increased.[45,46] Specific environmental factors that have been investigated include dietary compounds, including cow's milk, wheat gluten, soy products, fats, and even coffee or tea, along with vitamin deficiencies, *N*-nitroso compounds and other toxins, and viral infections.[44] One specific hypothesis associates the development of T1D in Finland with seasonal changes in climate, it being more common during the cold season.[47,48] Although the genetic argument has long been considered the main factor in triggering T1D, new studies are pointing to the importance of the environment.

VIRAL TRIGGERS OF TYPE 1 DIABETES

Although multiple environmental factors have been implicated as possible triggers of T1D, viral disease induction has long been argued as one of the main potential environmental triggers. Viruses in general have been receiving much attention as possible triggers in T1D because of their ability to mechanistically generate an active immune response when encountered in the host. In the case of T1D, the host environment is primed with immune cells poised to activate; all they need is an immunological stimulus. Researchers in the aforementioned seasonal pattern study on T1D induction even argue that the primary culprit responsible for triggering induction of disease during the winter season may actually be viral infections.[47,48] Several viruses have been directly implicated as potential triggers of T1D, including enterovirus, adenovirus, Coxsackie B virus, cytomegalovirus, hepatitis C virus, mumps virus, rotavirus, and rubella virus.[49–56] The presence of a viral infection can lead to immune cell activation through various possible mechanisms. Viruses may directly alter a host cell or tissue in such a manner that the immune system identifies it as an immunological target. Targeted host cells are lysed, releasing self peptides and fragments of the host cell into the circulation, where they may be processed and presented via APCs.[49,54] Viruses can directly alter the immune system of the host by inducing polyclonal B cell activation, the release of lymphokines, the activation of immune cells, and the disruption of the strict immune balance between T helper 1 and T helper 2 type responses, thus promoting unwarranted immune activation.[54] It has been hypothesized that antiviral antibodies also can lead to the formation of anti-idiotypic antibodies, which can become autoreactive if the first antibody is generated against the part of the virus that interacts with the host.[54]

Molecular mimicry is by far one of the most well studied processes associated with viral triggering mechanisms and T1D disease induction. Viruses produce proteins similar to those of the host. Although not all viral proteins share homology with the host, certain viral components share a distinct homology with identified β-cell antigens targeted in an autoimmune response.[44,54] Upon reacting to a viral infection, the immune system may process and present a homologous viral protein in such a manner that the epitope targeted by the immune system can interact with both self and viral proteins. This process becomes especially important with respect to T1D disease induction if the homologous viral protein that is recognized and processed shares a distinct homology with β-cell–associated proteins. Unintentional activation of an immune response geared toward β-cell–associated proteins put together with epitope spreading of antigens is a volatile combination for the host.

GAD is a well-defined example of an autoantigen in T1D that shares homology with viral proteins. Researchers have demonstrated that GAD peptides share mimicry with the P2-C viral sequence of the Coxsackie B virus and the major outer capsid protein of rotavirus.[55,56] Immunization studies of mice using homologous Coxsackie viral sequences have induced T cell immune responses that cross-react with GAD peptides, showing that viral homology may induce responses to self proteins.[57] Furthermore, researchers have linked rotavirus infection and pancreatic islet autoimmunity in children on the HLA-DR4 background, linking both viral induction and genetic susceptibility.[56] Researching molecular mimicry and viral induction of T1D provides a pivotal bridge between an environmental factor and an immunological system responsible for T1D. Whether T1D is triggered environmentally, genetically, or in combination, it is still not known how exactly to prevent disease progression once it starts. Several other unknown immunological factors may play a role in triggering T1D.

INNATE IMMUNE SYSTEM

The innate immune system is the body's first line of defense against invading microbes and pathogens. Although this system does not generate long-term antigen-specific immunological memory, this system does have a high degree of immunological specificity when encountering potential pathogens.[8] The innate immune system is composed of macrophages, dendritic cells, natural killer cells, neutrophils, and epithelial cells, all of which have their own unique role in the framework of an innate response. The role of the innate immune system is to survey and detect pathogens via host cell receptors that are able to recognize common structural elements expressed by pathogens. These pattern recognition receptors (PRRs), though specific, are not clonal and lack the ability for clonal expansion of cells, as seen in the adaptive immune system.[58–60]

Cells of the innate immune system use several types of PRRs to help initiate an immune response. Toll-like receptors (TLRs) are an evolutionarily conserved class of PRRs, which when activated cause the activation of the immune system.[58,60,61] These receptors recognize specific microbial membrane components, bacterial flagellin, and DNA and RNA from bacteria and viruses.[63] Upon recognition of a ligand, TLR activation brings about a cascade of proinflammatory cellular responses, including up-regulated production of cytokines, chemokines, and costimulatory molecules.[58,62–65] TLRs recognize numerous self-expressed mammalian molecules in addition to the known nonself molecules.[66] These self antigens are usually indicators of stress and disease or are molecules that are modified because of the disease.[66,67]

Once the innate immune system becomes activated, it in turn has the ability to promote activation of the adaptive immune system.[58,62–65] The activation of the innate immune system is a prerequisite for the initiation of specific adaptive immune responses, such as T-helper 1 type responses.[59,67,68] It has been suggested that in T1D and other autoimmune diseases, TLRs may be priming an unwarranted adaptive immune response because of autoreactive processes directed against self antigens.[69,70] Innate immune cells activated against self antigens may be responsible for a break in tolerance. Once tolerance is broken, the body may promote autoreactive immune responses.

TLRs are not the only PRRs used by cells of the innate immune system. Non-TLR PRRs include Nod-like Receptors (NLRs), Triggering Receptors Expressed on Myeloid Cells (TREMs), and C-type Lectin Receptors (CLRs) have not been directly linked to autoimmunity or T1D, but represent a new class of receptors that may be responsible for triggering autoimmune responses or autoimmune initiation of the adaptive immune system.[74–76] Any immunological trigger directed against self is potentially dangerous. Innate immune cells triggered via self ligand binding to TLRs or non-TLR PRRs could in turn activate T cells and an adaptive immune response because of the production of cytokines and other inflammatory signals. Once the balance of the immune system is flipped toward autoreactivity, negative outcomes are sure to follow.

The innate immune system has been well studied in the context of T1D development in both humans and the NOD mouse. T1D disease progression has been characterized in the 2 phase processes of asymptomatic inflammation of the islets followed by autoimmune destruction of the islets. During phase 1, macrophages infiltrate the pancreatic islets and may be responsible for the asymptomatic inflammation that takes place there.[8–13] Once activated, macrophages can secrete nitrogen and oxygen free radicals, as well as inflammatory cytokines, into their surrounding microenvironment.[74] These molecules can either directly damage surrounding cells or initiate cellular damage by in turn activating other cell types to cause damage. It has even been illustrated that without the presence of macrophages in NOD mice, differentiation of β-cell–cytotoxic T cells does not take place.[75–77] Restoring macrophages in these types of depletion experiments restores the ability of β-cell–cytotoxic T cell generation,[75–77] thus illustrating the importance of macrophages in T1D disease progression. The innate and adaptive immune systems communicate using an intricate system of checks and balances. Once one end of the system becomes disrupted, the whole system can falter.

ADAPTIVE IMMUNE SYSTEM

The adaptive immune system is the body's second line of defense against pathogens and disease, activated by the innate immune system. This system is antigen-specific, facilitating the generation of immunological memory, which serves to eliminate reoccurring pathogens more effectively upon repeat exposure to a given pathogen. This highly specific system uses receptor interaction between T cells and APCs to determine self from nonself. The hallmarks of an adaptive immune response are the generation of long-term immunological memory, peptide presentation in the context of MHCs, and the production of antigen-specific antibodies.[78]

T lymphocytes and their specific TCRs are a crucial part of the adaptive immune system. T lymphocytes are generated in the thymus, each having its own structurally diverse and unique antigen-recognizing receptor known as the TCR. The TCR recognizes processed antigenic peptides presented in the context of MHC via APCs. Once recognition of a peptide-bound MHC complex occurs via the TCR, a cascade of signaling events takes place, leading to either a CD4+ T cell or CD8+ T cell response, depending on which class of MHC, either MHC class I or MHC class II, was initially recognized by the TCR. Ultimately, once a TCR recognizes a processed peptide in the context of an MHC, immunological events take place that bring about either the direct targeted destruction of cells presenting the peptide or the generation of an antibody response and further T cell activation.

In phase 2 of T1D disease progression, both T and B lymphocytes accumulate in an islet lesion, where they can be activated against self antigens to trigger an immunological response toward the self. This autoreactive immune response then leads to the destruction of the insulin-producing β-cell within the pancreas. Both MHC class I and MHC class II restricted T cells are necessary for T1D disease progression in both humans and NOD mice.[79,80] Anti-insulin CD+4 T cell populations have been reported in NOD mice. It is thought that interactions between insulin peptides and MHC molecules can cause the development of an autoimmune T cell repertoire.[81] Although this process is not totally understood, it is necessary to determine what exact factors are causing the immune system to become unregulated in such a manner as to promote an autoimmune response.

REGULATION VIA T CELLS

Keeping the immune system tolerant is a strict balancing act: excessive regulation restricts proper immune function, whereas a lack of regulation permits overt immune function with the potential to target the self. Regulating when an immune response begins, continues, and ends is a very important part of the immune system's checks and balances. Regulatory T cells (T regs) and natural killer T (NKT) cells are 2 subtypes of T cells that are important regulators in the progression of T1D and autoimmune diseases.

T regs are a unique population of T cells that express the forkhead transcription factor forkhead box P3 (FOXP3) and are CD+4 CD+25.[91–93] T regs are well-known immunoregulators that can suppress proliferation of effector cells by shutting down IL-2 activation pathways.[85] T regs help the immune system maintain proper T cell homeostasis by preventing T cell activation, and this in turn leads to the development of inflammatory responses.[86,87] T regs work to shut off an active T cell response and are thought to prevent autoimmune development by regulating the expansion of T cell populations, T cell differentiation, and effector T cell function.[88] Although individuals with T1D have numbers of T regs equal to those of normal healthy individuals,[89–91] T regs from T1D individuals display decreased suppressive characteristics, which suggest a defect in T reg function in people suffering from T1D.[89,90] Adoptive transfer experiments using T regs in NOD mice have effectively been demonstrated to protect against the onset and progression of T1D.[92] Although current use of T regs in NOD mice shows promising therapeutic qualities, the potential use of T regs in humans is not sufficiently understood.

NKT cells share similar characteristics with both natural killer cells and T cells. These cells can cause direct cell lysis due to Fas-ligand interactions and induce cytotoxic damage of cells due to IFNg, like Natural Killer cells[102–104], but they are also similar to T cells in that they are biased toward TCR usage and produce IL-4.[102–104] When activated, NKT cells produce IL-4, which inhibits inflammatory T helper 1 type responses and promotes inhibitory T helper 2 type responses.[96,97] Individuals suffering from T1D show a decreased number of NKT cells along with a decreased ability to produce and secrete IL−4.[98] As seen with T regs, adoptive transfer experiments with NKT cells effectively protect against the onset and progression of T1D in NOD mice.[99] Because of these 2 types of T cells, individuals that cannot inhibit or properly regulate T cell responses face increased autoimmune problems.

CONCLUDING REMARKS

T1D results from autoimmune destruction/dysfunction of insulin-secreting cells. Under physiologic conditions, there is balance between pathogenic T cells that mediate disease such as T cells with marked conservation of their TCRs (eg, insulin) and regulatory cells that control autoimmunity. In T1D and other autoimmune disorders, there is an altered balance between pathogenic and T regs.

Autoantibodies are some of the most potent risk determinants for autoimmune diseases, with relative risk exceeding 100 when all 4 autoantibodies are present in an asymptomatic child. The archetypical model for the application of autoantibodies is T1D. Seminal studies have suggested that using a combination of humoral immunological markers gives a higher predictive value for T1D progression and greater sensitivity without a significant loss of specificity. There is a growing effort as well as a large opportunity for exploring novel strategies alone or in combination with immunomodulation with the ultimate goal of finding the cure for T1D.

Emerging evidence indicates that *in silico* research is complementary to current experimental approaches in T1D research and has the potential not only to assist researchers in designing laboratory experimentation but also to build a framework of data collection to predict the outcome of therapeutic strategies aimed at halting the β-cell–specific autoimmune process. With support from the Juvenile Diabetes Research Foundation, the Institute for Systems Biology has generated a T1D database (T1Dbase) with numerous data sets of molecular signatures of T1D (http://t1dbase.org/cgi-bin/dispatcher.cgi/welcome/display). Molecular signatures have revealed remarkable differentiation of organ-specific complications as reflected in the work of the Kretzler laboratory on transcription factor binding site patterns in nephropathy associated with T2DM versus nephropathy due to lupus erythematosus, glomerulonephritis, or other causes.[100,101] Similar studies on T1D-associated neuropathy, are being completed by our group at the National Center for Integrative Biomedical Informatics (https://portal.ncibi.org/portal). Mathematical modeling of the imbalance of T cell subsets in T1D, in work led by Patrick Nelson, is expected to inform our future work.

ACKNOWLEDGMENT

This work was supported by the National Institutes of Health (grants RO1 DK53456, DK56200, and NIDDK PA-04-081 to M.P.) and by the University of Michigan Center for Computational Medicine and Biology Pilot Research Program. The authors gratefully thank the Brehm Coalition for its support. Any research involving human subjects was conducted in accordance with the guidelines of the Declaration of Helsinki and was approved by the institutional review board; all subjects provided informed consent.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Janeway CA, Travis P, Walport M, Shlomchik M. *Immunobiology*. 6th ed. New York, NY: Garland Science; 2007.
2. Bardsley JK, Want LL. Overview of diabetes. *Crit Care Nurse Q* 2004; 27: 106–112.
3. Pietropaolo M, Barinas-Mitchell E, Kuller LH. Perspectives in diabetes heterogeneity of diabetes mellitus. Unraveling a dispute: is systemic inflammation related to islet autoimmunity? *Diabetes* 2007; 56: 1189–1197.
4. Delovitch TL, Singh B. The nonobese diabetic mouse as a model of autoimmune diabetes: immune dysregulation gets the NOD. *Immunity* 1997; 7: 727–738.
5. Katz JD, Benoist C, Mathis D. T helper cell subsets in insulin-dependent diabetes. *Science* 1995; 268: 1185–1188.
6. Rabinovitch A, Suarez-Pinzon WL, Sorensen O. Interleukin 12 mRNA expression in islets correlates with beta-cell destruction in NOD mice. *J Autoimmun* 1996; 9: 645–651.
7. Von Herrath MG, Oldstone MB. Interferon-gamma is essential for destruction of beta-cells and development of insulin-dependent diabetes mellitus. *J Exp Med* 1997; 185: 531–539.
8. Wong FS, Janeway CA. The role of CD4 and CD8 T cells in type 1 diabetes in the NOD mouse. *Res Immunol* 1997; 148: 327–332.
9. Yoon JW, Jun HS, Santamaria P. Cellular and molecular mechanisms for the initiation and progression of beta-cell destruction resulting from the collaboration between macrophages and T cells. *Autoimmunity* 1998; 127: 109–112.
10. Lee KU, Kim MK, Amano K, et al. Preferential infiltration of macrophages during early stages of insulitis in diabetes-prone BB rats. *Diabetes* 1989; 37: 1053–1058.
11. Voorbij HA, Jeucken PH, Kabel PJ, et al. Dendritic cells and scavenger macrophages in pancreatic islets of pre-diabetic BB rats. *Diabetes* 1989; 38: 1623–1629.
12. Jansen A, Homo-Delarche F, Hooijkaas H, et al. Immunohistochemical characterization of monocyte-macrophage and dendritic cells in the initiation of insulitis and beta cell destruction in NOD mice. *Diabetes* 1994; 43: 667–675.
13. Lee KU, Amano K, Yoon JW. Evidence for initial involvement of macrophages in development of insulitis in NOD mice. *Diabetes* 1988; 37: 989–991.

14. Kawasaki E, Gill RG, Eisenbarth GS. *Type 1 Diabetes Mellitus*. Austin, TX: Landes Bioscience; 1999.
15. Yoon JW, Jun HS. Cellular and molecular pathogenic mechanisms of insulin-dependent diabetes mellitus. *Ann N Y Acad Sci* 2001; 928: 200–211.
16. Pearl-Yafe M, Kaminitz A, Yolcu ES, et al. Pancreatic islets under attack: cellular and molecular effectors. *Curr Pharm Des* 2007; 13: 749–760.
17. Estella E, McKenzie MD, Catterall, et al. Granzyme B-mediated death of pancreatic beta-cells requires the proapoptotic BH-3only molecule bid. *Diabetes* 2006; 55: 2212–2219.
18. Kawasaki E, Abiru N, Eguchi K. Prevention of type 1 diabetes: from the view point of beta cell damage. *Diabetes Res Clin Pract* 2004; 55(suppl 1): S27–S32.
19. Ichinose K, Kawasaki E, Eguchi K. Recent advancement of understanding pathogenesis of type 1 diabetes and potential relevance to diabetic nephropathy. *Am J Nephrol* 2007; 27: 554–564.
20. Atkinson MA. Eisenbarth GS. Type I diabetes: new perspectives on disease pathogenesis and treatment. *Lancet* 2001; 358: 221–229.
21. Finegood DT, Scaglia L, Bonner-Weir S. Dynamics of β-cell mass in the growing rat pancreas. Estimation with a simple mathematical model. *Diabetes* 1995; 44: 249–256.
22. Nakayama M, Abiru N, Moriyama H, et al. Prime role for an insulin epitope in the development of type 1 diabetes in NOD mice. *Nature* 2005; 435: 220–223.
23. Hutton JC, Wenzlau JM, Juhl K, et al. The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. *PNAS* 2007; 104: 17 040–17 045.
24. Herrath MV, Sanda S, Herold K. Type 1 diabetes as a relapsing-remitting disease? *Nat Rev Immunol* 2007; 7: 988–994.
25. Pietropaolo M, Yu S, Libman IM, et al. Cytoplasmic islet cell antibodies remain valuable in defining the risk of progression to type 1 diabetes in subjects with other autoantibodies. *Pediatr Diabetes* 2005; 6: 184–192.
26. Nepom, GT. Class II antigens and disease susceptibility. *Annu Rev Med* 1995; 46: 17–25.
27. Serreze DV, Leiter EH. Molecular pathology of insulin dependent diabetes mellitutus. *Current Directions in Autoimmunity*. 2001; 4: 31–67.
28. Redondo MJ, Fain PR, Krischer P, et al. Expression of beta-cell autoimmunity does not differ between potential dizygotic twins of siblings of patients with type 1 diabetes. *J Autoimmun* 2004; 23: 275–279.
29. Thomson G. HLA DR antigens and susceptibility to insulin-dependent diabetes mellitus. *Am J Hum Genet* 1984; 36: 1309–1317.
30. Todd JA, Walker NM, Cooper JD, et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* 2007; 39: 857–864.
31. Powis SH, Mockridge I, Kelly A, et al. Polymorphism in a second ABC transporter gene located within the class II region of the human major histocompatibility complex. *Proc Natl Acad Sci U S A* 1992; 89: 1463–1467.
32. Colonna M, Bresnahan M, Bahram S, et al. Alleleic variants of the human putative peptide transporter involved in antigen processing. *Proc Natl Acad Sci U S A* 1992; 89: 3932–3936.
33. Germain RN. MHC-dependent antigen processing and peptide presentation: providing ligands for T lymphocyte activation. *Cell* 1994; 76: 287–299.
34. Rudert WA, Trucco M. Rapid detection of sequence variation using polymers of specific oligonucleotides. *Nucleic Acid Res* 1992; 5: 1146.
35. Todd JA, Bell JI, McDevitt HO. HLA-DQβgene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* 1987; 329: 599–604.
36. Morel PA, Dorman JS, Todd JA, et al. Aspartic acid at position 57 of the HLA-DQ beta chain protects against type I diabetes: a family study. *Proc Natl Acad Sci U S A* 1988; 85: 8111–8115.
37. Trucco M. To be or not to be ASP 57, that is the question. *Diabetes Care* 1992; 15: 705–715.
38. Aly TA, Ide A, Jahromi MM, et al. Extreme genetic risks for type 1A diabetes. *Proc Natl Acad Sci U S A* 2006; 103: 14 074–14 079.
39. Brown JH, Jardetzky T, Saper MA, et al. A hypothetical model of the foreign antigen binding site of class II histocompatibility molecules. *Nature* 1988; 332: 845–850.
40. Brown JH, Jardetzky TS, Gorga JC, et al. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 1993; 364: 33–39.
41. Jardetzky TS, Brown JH, Gorga JC, et al. Three-dimensional structure of a human class II histocompatibility molecule complexed with superantigen. *Nature* 1994; 368: 711–718.
42. Turco E, Manfras BJ, Ge L, et al. The x boxes from promoters of HLA class II B genes at different loci do not compete for nuclear protein-specific binding. *Immunogenetics* 1990; 32: 117–128.
43. Demotz S, Grey HM, Sette A. The minimal number of class II MHC-antigen complexes needed for T cell activation. *Science* 1990; 249: 1028–1030.
44. Akerblom HK, Knip M. Putative environmental factors in type 1 diabetes. *Diabetes Metab Rev* 1998; 14: 31–67.
45. Herman R, Knip M, Veijola R, et al. The FINDIANE Study Group: temporal changes in the frequencies of HLA genotypes in patients with type 1 diabetes: indication of an increased environmental pressure? *Diabetologia* 2003; 46: 420–425.
46. Gillespie KM, Bain SC, Barnett AH, et al. The rising incidence of childhood type 1 diabetes and reduced contribution of high-risk HLA haplotypes. *Lancet* 2004; 364: 1645–1647.
47. Kimpimaki T, Kupila A, Hamalaninen AM, et al. The first signs of B-cell autoimmunity appear in infancy in genetically susceptible children from the general population: the Finnish Type 1 Diabetes Prediction and Prevention Study. *J Clin Endocrinol Metab* 2001; 86: 4782–4788.
48. Laron Z. Lessons from recent epidemiological studies in type 1 childhood diabetes. *J Pediatr Endocrinol Metab* 1999; 12(suppl 3): 733–736.
49. Yoon JW. Role of viruses in the pathogenesis of IDDM. *Ann Med* 1991; 23: 437–445.
50. Varela-Calvino R, Peakman M. Enteroviruses and type 1 diabetes. *Diabetes Metab Res Rev* 2003; 19: 431–441.
51. Green J, Casabonne D, Newton R. Coxsackie B virus serology and type I diabetes mellitus: a systemic

review of published case-control studies. *Diabet Med* 2004; 21: 507–514.

52. Barbu AR, Akusjarvi G, Welsh N. Adenoviral-mediated transduction of human pancreatic islets: importance of adenoviral genome for cell viability and association with a deficient antiviral response. *Endocrinology* 2005; 146: 2406–2414.
53. Chen LK, Chou YC, Tsai ST, et al. Hepatitis C virus infection-related type 1 diabetes mellitus. *Diabet Med* 2005; 22: 340–343.
54. Jun HS, Yoon JW. A new look at viruses in type I diabetes. *Diabetes Metab Res Rev* 2002; 19: 8–31.
55. Honeyman MC, Coulson BS, Stone NL, et al. Association between rotavirus infection and pancreatic islet autoimmunity in children at risk of developing type 1 diabetes. *Diabetes* 2000; 49: 1319–1324.
56. Atkinson, MA, Bowman MA, Campbell L, et al. Cellular immunity to a determinant common to glutamic decarboxylase and Coxsackie virus in insulin-dependent diabetes. *J Clin Invest* 1994; 94: 2125–2129.
57. Titan J, Lehmann PV, Kaufman DL. T cell cross-reactivity between Coxsackie virus and glutamate decarboxylase is associated with a murine diabetes susceptibility allele. *J Exp Med* 1994; 180: 1979–1984.
58. Takeda K, Kaisho T, Akira S. Toll-like receptors. *Annu Rev Immunol* 2003; 21: 335–376.
59. Dabbagh K, Lewis DB. Toll-like receptors and T-helper-1/T-helper-2 responses. *Curr Opin Infect Dis* 2003; 16: 199.
60. Pasare C, Medzhitov R. Toll-like receptors and acquired immunity. *Semin Immunol* 2004; 16: 23–26.
61. Akira S, Uematsu S, Takeuchi O. Pathogen recognition and innate immunity. *Cell* 2006; 124: 783–801.
62. Godfrey DI, Hammond KJ, Poulton LD, et al. NKT cells: facts, functions and fallacies. *Immunol Today* 2000; 21: 573–583.
63. Medzhitov R. Toll-like receptors and innate immunity. *Nat Rev Immunol* 2001; 1: 135–145.
64. Akira S, Takeda K. Toll-like receptor signaling. *Nat Rev Immunol* 2004; 4: 499–511.
65. Kawai T, Akira S. TLR signaling. *Nature* 2006; 13: 816–825.
66. Johnson GB, Brunn GJ, Platt JL. Activation of mammalian Toll-like receptors by endogenous agonists. *Crit Rev Immunol* 2003; 23: 15.
67. Termeer C, Benedix F, Sleeman, J, et al. Oligosaccharides of hyaluronan activate dendritic cells via toll-like receptor 4. *J Exp Med* 2002; 195: 99.
68. Buttari B, Profumo E, Mattei V, et al. Oxidized beta2glycoprotein I induces human dendritic cell maturation and promotes a T helper type 1 response. *Blood* 2005; 106: 3880.
69. McSorley SJ, Ehst BD, Yu Y, Gewirtz AT. Bacterial flagellin is an effective adjuvant for CD4+ T cells in vivo. *J Immunol* 2002; 169: 3914–3919.
70. Martin DA, Elkon KB. Autoantibodies make a U-turn: the toll hypothesis for autoantibody specificity. *J Exp Med* 2005; 202: 1465–1469.
71. Fritz JH. Le Bourhis L Sellge G, et al. Nod-like proteins in immunity, inflammation and disease. *Nat Immunol* 2006; 12: 1250–1257.
72. Klesney-Tait J, Turnbull IR, Colonna M. The TREM receptor family and signal integration. *Nat Immunol* 2006; 12: 1266–1273.
73. Robinson MJ, Sancho D, Slack EC, et al. Myeloid C-type lectins in innate immunity. *Nat Immunol* 2006; 12: 1258–1265.
74. Beyan H, Buckley LR, Yousaf N, et al. A role for innate immunity in type 1 diabetes?. *Diabetes Metab Res Rev* 2003; 19: 89–100.
75. Jun HS, Yoon CS, Zbytnuik L, et al. The role of macrophages in T-cell mediated autoimmune diabetes in nonobese diabetic mice. *J Exp Med* 1999; 189: 347–358.
76. Oschilewski U, Kiesel U, Kolb H. Administration of silica prevents diabetes in BB rats. *Diabetes* 1985; 34: 197–199.
77. Lee KU, Pak CY, Amano K, Yoon JW. Prevention of lymphocytic thyroditis and insulitis in diabetes prone BB rats by the depletion of macrophages. *Diabetologia* 1988; 31: 400–402.
78. Medzhitov R, Janeway CA Jr. Innate immunity: the virtues of a nonclonal system of recognition. *Cell* 1997; 91: 295–298.
79. Bottazzo GF, Dean BM, McNally JM, et al. In situ characterization of autoimmune phenomena and expression of HLA molecules in the pancreas in diabetic insulitis. *New Engl J Med* 1985; 313: 353–360.
80. Conrad B, Weidman E, Trucco G, et al. Evidence for superantigen involvement in insulin-dependent diabetes mellitus etiology. *Nature* 1994; 371: 351–355.
81. Levisetti MG, Suri A, Petzold SJ, Unanue ER. The insulin-specific T cells of nonobese diabetic mice recognize a weak MHC-binding segment in more than one form. *J Immunol* 2007; 178: 6051–6057.
82. Khatrri R, Cox T, Yasayko SA, Ramsdell F. An essential role for scurfin in CD+4CD+25 T-regulatory cells. *Nat Immunol* 2003; 4: 337–342.
83. Fontenot JD, Gavin MA, Rudensky AY. FoxP3 programs the development and function of CD+4 CD+25 regulatory T-cells. *Nat Immunol* 2003; 4: 330–336.
84. Hori S, Nomura T, Sakaguchi S. Control of regulatory T cell development by the transcription factor Foxp3. *Science* 2004; 299: 1057–1061.
85. Randolph DA, Fathman CG. CD+4CD+25 regulatory T cells and their therapeutic potential. *Annu Rev Med* 2006; 57: 381–402.
86. St. Clair EW, Turka LA, Saxon A, et al. New reagents on the horizon for immune tolerance. *Annu Rev Med* 2007; 58: 329–346.
87. Bluestone JA, Tang Q. How do CD4+CD25+ regulatory T cells control autoimmunity? *Curr Opin Immunol* 2005; 17: 638–642.
88. Tang Q, Bluestone JA. Regulatory T-cell physiology and application to treat autoimmunity. *Immunol Rev* 2006; 212: 217–237.
89. Brusko TM, Wasserfall CH, Clare-Salzler MJ, et al. Functional defects and the influence of age on the frequency of CD+4CD+25 T-cells in type 1 diabetes. *Diabetes* 2005; 54: 1407–1414.
90. Lindley S, Dayan CM, Bishop A, et al. Defective suppressor function in CD+4CD+25 t cells from patients with type 1 diabetes. *Diabetes* 2005; 54: 92–99.

91. Putnam AL, Vendrame F, Dotta F, Gottlieb PA. CD+4 CD+25 high regulatory T cells in human autoimmune diabetes. *J Autoimmun* 2005; 24: 55–62.
92. Mason D, Powrie F. Control of immune pathology by regulatory T cells. *Curr Opin Immunol* 1998; 10: 649–655.
93. Bendelac A. Mouse NK+1 T cells. *Curr Opin Immunol* 1995; 7: 367.
94. Bendelac A, Rivera MN, Park SH, Roark JH. Mouse CD1-specific NK1 T cells–development, specificity, and function. *Ann Rev Immunol* 1997; 15: 535–562.
95. Macdonald HR. NK1.1+ T cell receptor-α β+ cell–new clues to their origin, specificity, and function. *J Exp Med* 1995; 182: 633–638.
96. Burdin N, Brossay L, Kronenberg M. Immunization with α-galactosylceramyde polarizes CD1-reactive NK T-cells toward Th2 cytokine synthesis. *Eur J Immunol* 1999; 29: 2014–2025.
97. Singh N, Hong S, Scherer DC, et al. Cutting edge: activation of NK T-cells by CD1d and α-galactosylceramyde directs conventional T cells to the acquisition of a Th2 phenotype. *J Immunol* 1999; 163: 237–2377.
98. Wilson SB, Kent SC, Patton KT, et al. Extreme Th1 bias of invariant Vα 24Jα Q T cells in type 1 diabetes. *Nature* 1998; 391: 177–181.
99. Hammond KJL, Poulton LD, Palimsano LJ, et al. α β-T cell receptor (TCR)+ CD-4 CD-8 (NKT) thymocytes prevent insulin-dependent diabetes mellitus in nonobese diabetic (NOD/Lt) mice by the influence of interleukin-4 (IL-4) and/or IL-10. *J Exp Med* 1998; 187: 1047–1056.
100. Smink LJ, Helton EM, Healy BC, et al. T1DBase, a community web-based resource for type 1 diabetes research. *Nucleic Acids Res* 2005; 1(33): D544–D549.
101. Cohen CD, Klingenhoff A, Boucherot A, et al. Comparative promoter analysis allows de novo identification of specialized cell junction-associated proteins. *PNAS* 2006; 103: 5682–5687.
102. Anderson MS, Bluestone JA. The NOD mouse: a model of immune dysregulation. *Annu Rev Immunol* 2005; 23: 447–485.

Complications of Type 1 Diabetes: New Molecular Findings

Alin Stirban, MD,[1] Peter Rösen, MD,[2] and Diethelm Tschoepe, MD[1]

[1] Diabetes Clinic, Heart and Diabetes Center North Rhine-Westphalia, Ruhr University Bochum, Bad Oeynhausen, Germany
[2] German Diabetes Center, Heinrich Heine University, Duesseldorf, Germany

ABSTRACT

Interventions targeting the treatment of diabetic complications have not been nearly as successful as initially estimated, despite a marked improvement in therapeutic options for diabetes. The need for understanding why some very promising interventions have failed demands a closer look at the pathomechanisms of the complications. Great strides have been made in understanding the pathology, and several important hypotheses have emerged in recent years. On this basis, Brownlee and coworkers suggested a unifying hypothesis integrating various mechanisms discussed in past years with an overproduction of reactive oxygen species as an initiating cause. This hypothesis and further hypotheses, as well as mechanisms, are highlighted in this article. The field of pathomechanisms of diabetic complications is very wide, and any attempt to completely cover it within a single article is unrealistic. Therefore, our purpose is to present the most relevant concepts underlying diabetic complications in an attempt to contribute to a better understanding and pinpoint areas that warrant further research. ***Mt Sinai J Med 75:328–351, 2008.*** © *2008 Mount Sinai School of Medicine*

Key Words: complications, diabetes, molecular.

The estimate that, by 2025, worldwide 5.4% (300 million) of adults will have diabetes[1] is alarming and highlights the importance of understanding and treating the disease and its complications. Although only 5% to 10% of people with diabetes have an autoimmune etiology, the complications have more time to develop and place a serious burden on the quality of life of those people and on the social and medical system.

Address Correspondence to:

Alin Stirban
Diabetes Clinic
Heart and Diabetes Center
North Rhine-Westphalia
Ruhr University Bochum
Bad Oeynhausen, Germany
Email: astirban@hdz-nrw.de

Hyperglycemia is the key event in type 1 diabetes and occurs because of damage to pancreatic β-cells. Although systemic hyperglycemia affects all organs and cell types, several cell types, such as capillary endothelial cells in the retina, mesangial cells in the renal glomerulus, and neurons and Schwann cells in peripheral nerves, seem to be more prone to hyperglycemia-induced pathology.[2] The explanation is that the cells which are damaged by hyperglycemia are those that cannot efficiently regulate the intracellular transport of glucose into cells under hyperglycemic conditions, and this results in markedly increased internal glucose concentrations.[3] Therefore, the explanation for what causes diabetes complications has been proposed to involve mechanisms occurring inside these cells, rather than outside.[2]

PATHOMECHANISMS

Four major pathomechanisms are considered to lead to the development of complications in diabetes: (1) the polyol pathway and increased polyol pathway flux,[4] (2) increased hexosamine pathway throughput and subsequent overmodification of proteins by *N*-acetylglucosamine, (3) hyperglycemia-induced activation of protein kinase C (PKC) isoforms, and (4) increased formation of advanced glycation end products (AGEs; Figure 1).

For a long time, these pathomechanisms have been regarded as independent processes, or at least their interrelationship remained unclear. It was the intriguing idea of Brownlee[2] to suggest that these processes are linked together and have a common

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20057

© 2008 Mount Sinai School of Medicine

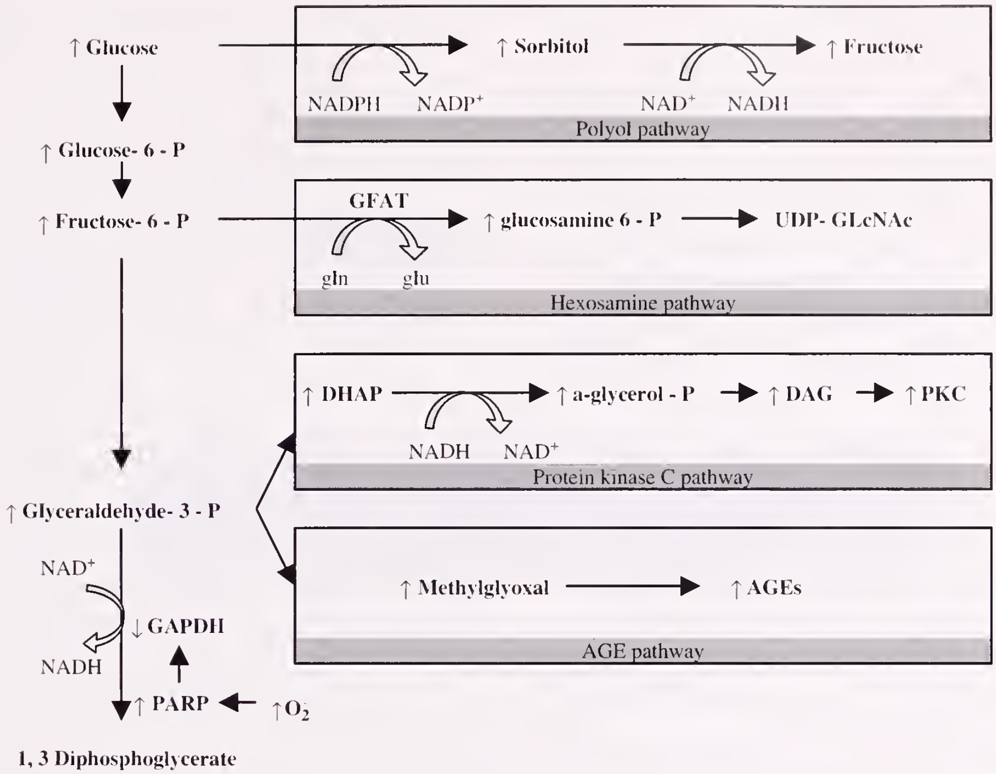


Fig 1. Inhibition of GAPDH by superoxides potentiates the 4 major pathomechanisms of hyperglycemia by an accumulation of intermediate products of glycolysis. **Abbreviations:** AGE, advanced glycation end product; DAG, diacylglycerol; DHAP, dihydroxyacetone phosphate; GAPDH, glyceraldehyde-3 phosphate dehydrogenase; GLcNAc, *N*-acetylglucosamine; NAD$^+$, nicotinamide adenine dinucleotide; NADH, reduced nicotinamide adenine dinucleotide; NADP$^+$, nicotinamide adenine dinucleotide phosphate; NADPH, reduced nicotinamide adenine dinucleotide phosphate; PARP, poly(adenylribose)polymerase; UDP, uridine diphosphate. Reprinted with permission from *Nature*.[11] Copyright 2001, Nature Publishing Group.

hyperglycemia (and also fatty acid) -dependent cause: the overproduction of superoxide by the mitochondrial electron transport chain resulting in oxidative stress (OS; Figure 2). Nowadays, we assume that OS plays not only a role in the induction of complications in diabetes but also in the destruction of β-cells leading to diabetes, the induction of insulin resistance, and the development of cardiovascular disease.[5–7]

Reactive oxygen species (ROS), such as superoxide and hydrogen peroxide, have traditionally been thought to be toxic byproducts of cellular metabolism that nonspecifically damage nucleic acids, proteins, lipids, and other cellular components. There is now a large body of evidence showing that ROS at moderate concentrations act as signaling molecules and play an important role in the regulation of various vascular cell functions.[8–12] In vascular endothelial cells, it has also been shown that ROS regulate vascular tone, oxygen sensing, cell growth and proliferation, apoptosis, and inflammatory responses. In addition to these regulatory functions under physiological conditions, excessive or sustained generation of ROS has been implicated in the pathogenesis of various cardiovascular diseases, such as atherosclerosis, hypertension, and diabetes and its complications.[11,13]

That the production of free radicals is increased in diabetic patients seems to be generally accepted. Several clinical studies have shown increases in levels of OS markers, such as ROOH, 8-epi-prostaglandin F2 alpha, 8-hydroxydeoxyguanosine, and oxidized low-density lipoprotein (LDL), in patients with insulin-dependent diabetes mellitus or non–insulin-dependent diabetes mellitus in comparison with healthy age-matched subjects.[7,14–17] Dandona and coworkers[14] showed an approximately 4-fold higher median concentration of 8-hydroxydeoxyguanosine in mononuclear cells of diabetic patients compared to corresponding controls. This difference was statistically significant and demonstrated for the first time greater oxidative damage to DNA in diabetic patients in comparison with controls. Clinical studies by Nourooz-Zadeh *et al.*[15,18,19] showed an approximately 2-fold increase in the plasma OS

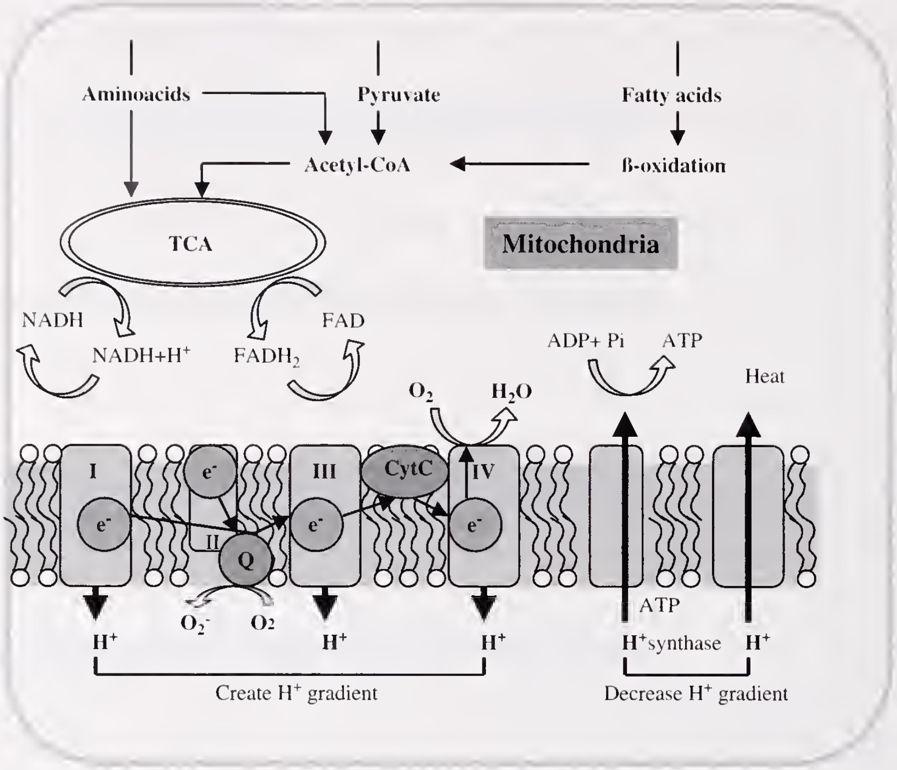


Fig 2. Mitochondrial electron transport chain, ATP, and superoxide production. Mitochondrial membrane complexes are termed I, II, III, and IV. **Abbreviations:** ADP, adenosine diphosphate; ATP, adenosine triphosphate; CoA, coenzyme A; CytC, cytochrome C; FAD, flavin adenine dinucleotide; FADH$_2$, reduced flavin adenine dinucleotide; GFAT, glutamine fructose-6 phosphate amidotransferase; NADH, reduced nicotinamide adenine dinucleotide; Pi, inorganic phosphorus; PKC, protein kinase C; Q, coenzyme Q; TCA, tricarboxylic acid. Adapted from *Diabetes*.[2]

measures 8-epi-prostaglandin F2 alpha and ROOH in diabetic patients in comparison with healthy subjects. It is important to note that the generation of ROS is elevated not only in the diabetic state but also in prediabetic stages, and this suggests a role of ROS in the process of diabetes manifestation.[20–23]

Although the overproduction of ROS in diabetes is generally accepted, the underlying mechanisms are a matter of discussion. In addition to autoxidation of glucose, a variety of cellular enzyme systems have been suggested as potential sources of ROS, including NADPH oxidase, xanthine oxidase, uncoupled endothelial nitric oxide synthase (eNOS), arachidonic acid–metabolizing enzymes including cytochrome *P*-450 enzymes, lipoxygenase, and cyclooxygenase. In addition, the mitochondrial respiratory chain may be the main source of ROS.[13,24] However, we have to take into account that the source may vary with the tissues and cells involved. In vascular cells, 4 enzyme systems seem to predominate: NADPH oxidase, xanthine oxidase, uncoupled eNOS, and mitochondrial electron leakage.[7,11,13,24,25] The functional significance of mitochondria-derived ROS in vascular endothelial cells has received little attention for a long time because vascular cells generally exhibit low metabolic activity[26] and mitochondria-generated ROS seem to be less well regulated than other systems such as NADPH oxidase.[27] Lately, the mitochondrial respiratory chain has gained much attention as a major source of ROS in most mammalian cells and predominantly in vascular cells, and excess production of ROS from mitochondria has been implicated in aging, a range of degenerative diseases, and the manifestation of diabetes and its complications.[11,28–30] Therefore, the underlying mechanism is outlined in more detail.

The generation of ATP by ATP synthase within the mitochondria requires energy that is obtained from the voltage gradient across the mitochondrial membrane.[31,32] The voltage gradient is created by the mitochondrial electron transport chain, which consists of 4 protein complexes (termed I-IV). During the metabolization of glucose and its glycolytic end products, electron donors are generated. Reduced nicotinamide adenine dinucleotide, the main electron donor, and reduced flavin adenine dinucleotide (formed by succinate dehydrogenase) donate electrons to complexes I and II, respectively, and further to coenzyme Q. From here, electrons are successively transferred to complex III, cytochrome C, complex IV, and finally molecular oxygen, which is reduced to water. The energy resulting from the voyage of electrons within the mitochondrial electron transport chain is used for the proton pump across the mitochondrial membrane, resulting in the aforementioned voltage gradient used for ATP synthesis[2] (Figure 2).

Hyperglycemia causes overproduction of electrons and an increase in the voltage gradient. Once a critical threshold is reached,[33] electrons will escape from the electron transport chain and be transferred to molecular oxygen, resulting in the production of superoxide anions. These molecules are normally degraded by the mitochondrial superoxide dismutase and catalase to H_2O and O_2.[2] However, ROS accumulate if the production of superoxide anions exceeds the capacity of superoxide dismutase or if the capacity of degradation is diminished as in diabetes.[5]

Not only has increased production of ROS been observed as a result of enhanced glycolytic conversion of glucose, but similarly, electron donors (reduced nicotinamide adenine dinucleotide and reduced flavin adenine dinucleotide), derived from β-oxidation of fatty acids, cause mitochondrial overproduction of ROS by similar mechanisms and may potentiate the pathogenicity of high glucose concentrations.[2]

OS might result not only from an enhanced generation of ROS but also from diminution of the antioxidative capacity in plasma and within the cells. In healthy cells, the action of ROS is tightly restricted with respect to the site and duration of action. A decrease in antioxidant capacity has been observed in the plasma of diabetic patients by several research groups.[7,17] In a prospective study, Salonen *et al.*[34] found a decline of plasma RRR-α-tocopherol levels after the onset of type II diabetes. Thornally *et al.*[35] reported a decrease of the endogenous antioxidant glutathione in the erythrocytes of diabetic patients. This decline in plasma levels of antioxidants such as RRR-α-tocopherol or glutathione may be caused by an increased production of free radicals in diabetic patients but also may reflect an impaired nutritional antioxidant intake.[36]

Despite the overall convincing evidence for an enhanced generation of ROS and a diminution of the antioxidative capacity in diabetes and prediabetic stages, the causal relationship between OS and the manifestation of diabetes and its complications remains a matter of discussion. Experimental studies in mice and rats have clearly demonstrated a causal relationship. High concentrations of antioxidants (vitamin E, vitamin C, *N*-acetylcysteine, and lipoic acid) or an overexpression of enzymes with antioxidative activity, such as superoxide dismutase and catalase, prevents or at least delays the manifestation of diabetes as well as the development of vascular complications.[7,11,17,37,38]

However, in most clinical studies, the treatment of diabetic patients with antioxidants has failed to prevent or improve diabetes or the development of its complications.[17,39,40] The reason for this inefficiency of antioxidant treatment in clinical settings is not yet completely clear; besides the dosage, one reason may be the low specificity of the antioxidant used. Most do not directly target the various sources of ROS generation but interact in a more or less unspecific manner with the complex network of antioxidant activities. In addition, it has to be taken into account that the antioxidant potency of different cells varies and that cells such as vascular cells and also β-cells are more sensitive to OS; these differences are not taken into account if one antioxidant or even a mixture of antioxidants is applied to patients.

The increased OS has 2 immediate consequences: the activation of nuclear factor κB (NF-κB) and the activation of PARP.

NF-κB is an important transcription factor leading to the activation of several proadhesive and proinflammatory pathways. It can also induce apoptosis (loss of vascular, muscular, and β-cells). It impairs the stability of progenitor cells, which are a relevant component for various repair mechanisms. It further enhances the generation of AGEs, which lead to a further amplification of the deleterious OS mechanisms. Extensive reviews on the effects and regulation of NF-κB exist.[41,42] Briefly, NF-κB resides in the cytoplasm of resting cells, being bound in its inactive form to the inhibitor molecule IκBα.[42] Upon activation, IκBα is rapidly phosphorylated and degraded, releasing NF-κB that will be translocated into the nucleus.[42] NF-κB binds here to DNA sequences and activates the transcription of NF-κB–regulated target genes, encoding substances such as cytokines, adhesion molecules, prothrombotic and vasoconstrictive gene products, receptor for advanced glycation end products (RAGE) itself, and IκBα.[42,43] A characteristic of RAGE-mediated activation of NF-κB is the long duration that exceeds inhibitory signals from autoregulatory loops.[41] Interestingly, a number of antiapoptotic genes, including Bcl-XL, Bcl-2, and the Bcl-2 homologue A1, are also regulated by NF-κB.[41,42] Therefore, NF-κB activation can also promote cellular survival in the absence of new protein synthesis by generating a rapid and sensitive cellular response.[41]

ROS also provoke activation of poly(adenyl ribose)polymerase, which not only causes a breakdown of reduced NADH and depletion of ATP but also has been shown to ribosylate glycerolaldehyde dehydrogenase, leading to an inhibition of this rate-limiting enzyme and thereby to an inhibition of glycolysis.[44,45] As a result, C3 and C6 intermediates of glucose conversion accumulate, enhancing the formation of diacylglycerol, which is a potent activator of protein kinase activity. In addition, an accumulation of C6 intermediates of glucose favors the polyol pathway and the formation of hexosamine phosphate, enabling the O-glycosylation of various proteins such as nitric oxide synthase (NOS) and the transcription factor SP1.[46,47] Thus, the increased formation of ROS activates a complex network of pathways involved in the development of vascular complications and also in the destruction of β-cells: PKC, the polyol and hexosamine pathways, and the activation of NF-κB. The central predication of the unifying mechanism proposed by Brownlee and coworkers[11,12] suggests that an increase in ROS is sufficient to induce the complex pattern of reactions leading finally to diabetes and its complications.

The Polyol Pathway

The enzyme aldose reductase is the key enzyme of the polyol pathway, normally reducing toxic aldehydes within the cells to inactive alcohols. When hyperglycemia occurs, aldose reductase also reduces glucose to sorbitol, which is later oxidized to fructose. This process consumes reduced NADPH,[48] an essential cofactor for regenerating reduced glutathione, an important intracellular antioxidant. By reducing the amounts of reduced glutathione, the polyol pathway increases susceptibility to intracellular OS.[2]

Indeed, it has been demonstrated in type 1 diabetic patients with nephropathy that antioxidant resources are scarce, whereas no such changes are present in people with uncomplicated diabetes or nephropathy from other causes.[49]

A cross-sectional study in young diabetic patients showed that systemic OS is present upon early onset of type 1 diabetes and further increased by early adulthood, and it suggested that appropriate support for enhancing the antioxidant supply in these young diabetic patients may help prevent clinical complications during the course of the disease.[20]

Beyond the impairment of antioxidant defense, excessive activation of the polyol pathway was previously suggested to determine intracellular accumulation of sorbitol, causing osmotic damage to vascular cells.[50] Although this hypothesis has lost importance lately, decreased nicotinamide adenine dinucleotide/NADPH and adenine nucleotide phosphorylation are probably the 2 main pathomechanisms that can be partly reversed by an aldose reductase inhibitor.[51]

Increased Hexosamine Pathway Activity

Most glucose is metabolized under normoglycemic conditions via glycolysis, and the first 2 products are glucose-6 phosphate and fructose-6 phosphate. However, during intracellular hyperglycemia, some of fructose-6-phosphate is diverted into a signaling pathway in which the enzyme glutamine fructose-6 phosphate amidotransferase converts it to glucosamine-6 phosphate and finally to uridine diphosphate *N*-acetylglucosamine.[2] *N*-Acetylglucosamine reacts with serine and threonine residues of transcription factors, resulting in pathologic changes in gene expression.[47] Sequelae include increased transcription of transforming growth factor-β1 (TGF-β1) and plasminogen activator inhibitor-1.[46]

Protein Kinase C Activation

Intracellular hyperglycemia increases diacylglycerol synthesis, a critical activating cofactor for the 3 classic isoforms of PKC: PKC-β, PKC-δ, and PKC-α.[23–26] In particular, the PKC-β isoform is strongly implicated in the pathogenic processes involved in diabetic microangiopathy, such as ischemia, vascular leakage, neovascularization, and abnormal vasodilator function.[52]

Activated PKC has a variety of effects on gene expression that lead, for example, to decreased antiatherosclerotic factors (eg, decreased activity of eNOS[53]) and increased proatherogenic factors (eg, endothelin-1, TGF-β, and plasminogen activator inhibitor-1).[52] In endothelial cells, PKC also activates NF-κB.[54]

Increased Formation of Advanced Glycation End Products and Advanced Glycation End Product–Mediated Signaling

Intracellular production of AGE precursors appears to damage cells by at least 4 mechanisms: (1) intracellular proteins, including proteins involved in the regulation of gene transcription, are modified[55]; (2) AGE precursors leave the cells by diffusion and modify extracellular matrix molecules in the vicinity,[56] thus changing signaling between the matrix and the cells and causing cellular dysfunction[57]; (3) AGE precursors, escaping from the intracellular environment, modify circulating proteins in the bloodstream (eg, albumin), thus altering their function; and finally (4) circulating proteins modified by AGE (of intracellular or extracellular origin) can bind to and activate AGE receptors, thereby altering the production of inflammatory cytokines and growth factors, which in turn are responsible for cellular and tissue damage.[58,59]

AGE receptors are present on the surface of different cells such as macrophages, adipocytes, endothelial cells, and vascular smooth muscle cells.[60–63] Several types of receptors have been described, such as scavenger receptors (macrophage scavenger receptor-AI, macrophage scavenger receptor-AII, CD68, and CD36), RAGE, AGE-R1, AGE-R2, and AGE-R3.[64–66] Although scavenger receptors are responsible for the removal of AGEs, RAGE seems to mediate most of the biological effects of AGEs.[67] RAGE is a member of the immunoglobulin multiligand receptor family involved in intracellular signal transduction.[41] Its activation promotes inflammatory response, apoptosis, prothrombotic activity, expression of adhesion molecules, and OS.[7,41,68]

Through the interaction between AGE and RAGE, a signaling cascade partly mediated by OS within the cell is activated, resulting also in the activation of NF-κB, the consequences of which have already been briefly discussed. However, recent data have brought into question whether carboxymethyllysine, a major AGE, can bind to RAGE,[69] thus suggesting that not all AGEs activate RAGE.

Another enzyme with harmful potential is inducible nitric oxide synthase (iNOS), which resides mainly in inflammatory cells and underlies the regulation by inflammatory cytokines.[70] iNOS can also be stimulated by OS in an NF-κB–dependent manner,[71] resulting in the production of high, toxic concentrations of nitric oxide (NO). The latter reacts with oxygen radicals, forming a highly reactive metabolite, peroxynitrite, which interacts with proteins and DNA, causing nitration of proteins (nitrative stress), DNA damage, further activation of NF-κB, caspase-3, and vascular cell apoptosis[72,73] (Figure 3).

RAGE is up-regulated in atheromatous lesions[74] and in diabetes,[75] whereas in the latter condition, the AGE scavenger receptors are down-regulated; this mechanism is probably also responsible for reduced AGE clearance.[76,77] Because AGE production is increased in diabetes and RAGE is overexpressed, it is conceivable that AGE-mediated pathomechanisms are exacerbated in this disease.[78]

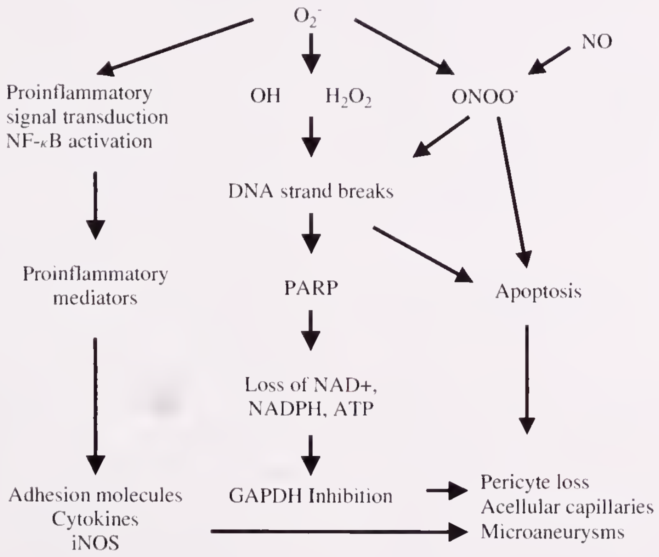

O_2^-
NO
Proinflammatory signal transduction NF-κB activation
OH
H_2O_2
ONOO⁻
DNA strand breaks
Proinflammatory mediators
PARP
Apoptosis
Loss of NAD+, NADPH, ATP
Adhesion molecules Cytokines iNOS
GAPDH Inhibition
Pericyte loss Acellular capillaries Microaneurysms

Fig 3. Pathomechanisms of diabetic retinopathy. **Abbreviations:** ATP, adenosine triphosphate; GAPDH, glyceraldehyde-3 phosphate dehydrogenase; iNOS, inducible nitric oxide synthase; NAD^+, nicotinamide adenine dinucleotide; NADPH, reduced nicotinamide adenine dinucleotide phosphate; NF-κB, nuclear factor κB; PARP, poly(adenylribose)polymerase.

One further characteristic of diabetes is the activation of inflammation. Rather than being a mechanism of pathology, the inflammation is a result of cellular activation and cytokine secretion subsequent to pathway activation. Recent data have suggested that there are several common pathogenic mechanisms underlying type 1 diabetes and diabetic nephropathy, such as susceptibility genes and proinflammatory cytokines that promote cell infiltration, proinflammatory signaling, and apoptosis.[79] Exacerbation of inflammation might be responsible not only for β-cell destruction within the pancreas but also for the production of oxygen reactive species, proinflammatory cytokines, metalloproteinase, and growth factors that are implicated in different stages of diabetic nephropathy.[79] Monocyte and macrophage accumulation has been identified in diabetic nephropathy, and this suggests a chronic inflammatory state.[79] However, the chronic inflammation encountered in diabetes contributes also to the development of further diabetes complications such as angiopathy, neuropathy, and retinopathy.

DIABETES COMPLICATIONS

Macroangiopathy

Atherosclerosis is a process that begins very early in life. Fatty streaks have been identified in the human aorta in the first decade of life and in the coronary arteries in the second decade.[80] The process evolves discontinuously over decades, with periods of stagnation and exacerbation. One characteristic of atherosclerosis is that lesions occur focally. An increase in the prevalence of cardiovascular disease in people with diabetes versus healthy individuals already exists in people 30 years old (10% versus 2% in nondiabetics) and further increases over the next decade (40% versus 5% respectively).[81] Therefore, the existence of precipitating factors for the development of atherosclerosis in this patient group has to be hypothesized.

The response-to-injury theory states that endothelial dysfunction is the first step toward atherosclerosis.[82] Endothelial dysfunction is characterized by a diminished production of endothelial NO, a key mediator contributing to vascular health. Thus, the vascular-protective effects of NO, such as vasodilatation and its antiatherosclerotic, antioxidant, and antithrombotic properties, are impaired.[83] Such a defect is of special relevance in large vessels because there the regulation of vascular tone is mainly determined by NO, whereas in small vessels prostaglandins and other endothelial derived factors are the main regulators.[84] Endothelial NO results from the transformation of L-arginine into citrulline and NO by the enzyme NOS. Constitutive NOS has been described in endothelial (eNOS) and neuronal cells (neuronal NOS), and iNOS has been previously mentioned.

Numerous risk factors, which have been reviewed elsewhere,[81] may contribute to the development of endothelial dysfunction in diabetes. The most important ones are hyperglycemia, hyperlipidemia, hyperinsulinemia, and accelerated aging.[83] Particularly in diabetes, the increased generation of ROS is of special importance[5] because ROS can directly react with NO into peroxynitrite. In this way, not only is the concentration of bioactive NO diminished, but also the reactivity of eNOS is reduced, and another highly reactive oxygen species (peroxynitrite) results. Both processes promote vasoconstriction, increase lipid peroxidation and nitrative stress, and trigger inflammatory responses.

Shear stress (ie, the friction effect of the laminary blood flow on the vessel surface) has an important physiological role: by stimulating flow-sensitive surface proteins of endothelial cells, it activates intracellular cascades that result in increased NO production.[85] However, exacerbated shear stress induces inflammation and ROS production by a mechanism related to the activation of NF-κB.[86] It also increases vascular permeability and alters cytoskeleton integrity,[87] thus explaining the preference of atherosclerotic lesions forming regions of branching and geometric irregularities where flow is disturbed and the shear stress is therefore exacerbated. In people with diabetes, increased shear stress is also generated by impaired blood rheology due to increased blood pressure, impaired myogenic tone,[88] and abnormal Windkessel physiology due to arterial calcification.[89]

The accumulation of LDL in the subendothelial matrix is one of the triggering events of atherosclerosis and is promoted by increased permeability and leakage of the endothelial cell layer, a common finding in diabetes-induced endothelial dysfunction.[90] There are several particularities of LDL in diabetes that increase its atherogenicity. Oxidized LDL delivers cholesterol to the atherosclerotic plaque by macrophage uptake,[91] and diabetes is characterized by increased oxidized LDL.[92,93] The explanation is the abundance of ROS in this disease, but there is also reason to believe that LDL particles in diabetic subjects are more prone to oxidation. Indeed, glycation of LDL particles increases in parallel with hyperglycemia, potentiating the sensitivity of LDL to oxidation.[92] The glycation seems to also increase the non–receptor-mediated uptake of LDL particles.[94]

One further pathomechanism attributed to LDL is the generation of LDL antibodies, but this mechanism is still controversial.[91]

High-density lipoprotein (HDL) is a protective factor against atherosclerosis.[95] During the acute Phase of inflammation and in diabetes, ROS are increasingly generated. They can oxidize phospholipids and sterols, which in turn reduce the anti-inflammatory and antioxidative properties of HDL.[96] The reduction in oxidative and inflammatory pressure with, for example, apolipoprotein A-I mimetic peptides[97] restores the protective properties of HDL.[96] The effect of HDL in reducing LDL lipid peroxidation is more persistent than that of antioxidant vitamins and could therefore be more protective.[98] The antioxidant activity of HDL is largely due to the paraoxonase 1 (PON-1) located on it, and animal experiments have suggested that PON-1 protects against atherogenesis. However, there is evidence that PON-1 is reduced in diabetic patients.[99–101]

It has also be shown that hyperglycemia increases vascular permeability by phosphorylation and relaxation of specific cytoskeletal and adhesion proteins,[102] decreases cell-to-cell contact, and augments albumin permeability[103]; these are important processes that may contribute to the aggressive development of atherosclerotic lesions in diabetes.

An important feature of vascular disease is also composed of 2 calcifying processes: (1) atherosclerotic calcification, intimally oriented, eccentric, and initiating at the base of necrotic plaques via apoptotic vesicles originating from dead vascular smooth muscle cells,[104,105] and (2) concentric medial artery calcification.

Medial artery calcification represents a characteristic change occurring in diabetes and renal disease.[106,107] It has been excellently reviewed by Shao and coauthors[89] and is considered only briefly here. The first step consists of the deposition of apatitic calcium phosphate within the medial layer around mineralizing matrix vesicles released by vascular smooth muscle cells and calcifying vascular cells.[108,109] The process has several similarities to the mineralization of membranous bones.[108,110] The primordial signal seems to be driven by mural hyperglycemia, inflammatory cytokines [eg, tumor necrosis factor-α (TNF-α)], and OS.[111,112] They activate an aortic bone morphogenic protein-2 (BMP2) muscle homolog (Msh homeobox 2) signaling cascade within the adventitia.[89] These calcifying signals are then transferred to the medial layer via the vasa vasorum[111–114]; this process is potentiated by enhanced mural neoangiogenesis in diabetes.[115,116]

On the basis of the implicated cells, calcifying signals seem to originate (1) from fibrofatty adipose tissue that expands in the periaortic adventitia[111,113–115] and aortic valve interstitium,[117] releasing inflammatory cytokines, (2) from hyperglycemia-activated BMP2 production within pericytes and mesangial myofibroblasts,[111,118] and (3) from endothelial cells that produce BMP2 when activated by TNF-α, ROS, and shear stress.[119] The signals are then transferred to at least 2 cell types that have been proposed to produce mineralizing matrix vesicles, the cause of vascular calcification: (1) multipotent vascular mesenchymal progenitors that differentiate into calcifying vascular cells during vascular injury and (2) vascular smooth muscle cells that can undergo osteogenic or chondrogenic differentiation in response to hyperphosphatemia,[89] an important mechanism that triggers calcification, especially in chronic kidney disease.[120] Calcifying vascular cells are a macrovascular myofibroblast subtype with similarities to the microvascular pericyte[121] and have been found to act also as osteoprogenitors in vitro.[121]

Several inhibitors of mineralization have been described, such as fetuin, which mainly controls the metabolism of vascular matrix vesicles,[122] and osteopontin, which has more complex functions.[123] Their deficiency constitutes the other face of enhanced vascular mineralization.[89]

The initial signals for atherosclerotic calcification are inflammatory and redox-dependent.[124] BMP2 also seems to play an important role here because atherosclerotic calcification[125] is also driven by calcifying vascular cells and accompanied by collagen type I and type II deposition and angiogenic invasion.[126] These features explain further particularities of atherosclerotic calcification versus medial artery calcification, such as fibrosis, apoptotic body formation, and cholesterol crystal accumulation.[89] Exact mechanisms of diabetes-promoted atherosclerotic lesions have been reviewed elsewhere in detail.[81,104]

Microangiopathy

Microvascular changes play an important role in the pathogenesis of most of the organ complications in diabetes (eg, nephropathy, retinopathy, neuropathy, and diabetic foot syndrome). Although they share several pathomechanisms with macrovascular complications, they present several particularities (eg, permeability changes)[127] and alterations that are specific to each organ (eg, retina, nerves, and kidney). An important difference has been shown in clinical studies, which have revealed so far that the improvement of glycemic control leads in the long term to a reduced rate of microvascular complications, whereas the macrovascular risk is not diminished.[128]

Thus, the general processes already discussed (OS and the network of reactions induced by OS) may be very similar in both types of vessels, but some particularities seem to make the microvasculature more sensitive to OS, on which this review is focused. We start with some general remarks on mechanisms of microvascular complications and then proceed to organ-specific features.

Changes in vascular permeability during the course of diabetes have received much attention lately.[90] It seems that hyperglycemia can selectively alter the glycocalix structure, thus increasing vessel permeability to water and not to proteins and thus inducing edema. Activation of PKC seems to be of special importance in the processes promoting changes in vascular permeability, basement membrane thickening, extracellular matrix expansion, abnormal angiogenesis, and excessive apoptosis. An extensive review on this topic is available.[129] AGEs and amadori-glycated proteins might also alter vascular function by stimulating PKC and activating TGF-β and the expression of extracellular matrix proteins.[130]

Diabetic Retinopathy

Diabetic retinopathy is the leading cause of acquired blindness in adults. It is duration-dependent, with 50% of people with diabetes having retinopathy after 10 years, a number that increases to 90% after 25 years of diabetes.[131]

Biochemical abnormalities identified in diabetic retinas and associated with hyperglycemia include activation of PKC,[132] nonenzymatic glycation,[133] polyol formation,[134] increased hexosamine syn thesis,[135] and activation of growth factors that promote apoptosis[136] as opposed to those that promote survival. The production of excess ROS has been suggested to induce many of these abnormalities[137] (Figure 3).

It has been shown that diabetes causes a profound increase in retinal intracellular glucose exceeding that in the brain, and this testifies to the increased potential to harm. The accumulation of intracellular glucose is accentuated in these cells by the fact that the rate of glycolysis is not proportionally accelerated with the increasing concentrations of glucose and additionally decreases over time (eg, 3 weeks to 3 months[137]).

The hallmarks of incipient diabetic retinopathy are increased vascular permeability (despite thickening of the basement membrane) and progressive vascular occlusion.[138]

Retinal capillaries consist of endothelial cells, the basement membrane, and finally pericytes situated within the basement membrane and almost as numerous as endothelial cells. The ratio of pericytes to endothelial cells changes from 1:1 under normal conditions to 1:4 in diabetes.[139] Although pericyte loss is one of the first changes that accompany retinopathy development, the exact mechanisms are not known and are believed to be due either to an accumulation of toxic compounds[140,141] or to an active elimination process.[142] Pericytes are required for the development of a proper retinal vascular network in the forming retina. Although they inhibit endothelial cell growth[143] under normoglycemic conditions, they appear to have a protective effect during hyperglycemia when they increase repair capacity. In the adult retina, pericytes are therefore believed to promote survival of retinal endothelial cells. Pericyte loss is accompanied in some capillaries by randomly increased proliferation of endothelial cells, whereas apoptosis of endothelial cells and the formation of acellular capillaries are triggered in others.[144] The exact mechanisms are not known, but the importance of pericytes in the modulation of endothelial cells has been repetitively highlighted.

As a result of the development of acellular capillaries, a progressive vascular occlusion occurs, leading to hypoxia. In an attempt at neovascularization, ineffective capillaries are formed, leading to microaneurysms, another clinical feature of incipient diabetic retinopathy.[138] At this level, vascular permeability is increased and can lead to macular edema and hard exudates. Continuing this process, new blood vessels are formed that enter the vitreous humor and increase the risk of hemorrhage.

Not only pericytes control endothelial cell replication, but endothelial cells themselves are modulating the recruitment and elimination of pericytes.[138] Hyperglycemia is believed to be the central mechanism that triggers the pathological changes of diabetic retinopathy. It impairs cellular metabolism, inducing cellular dysfunction, and alters the cross-talk between pericytes and endothelial cells, leading to aberrant cell proliferation or apoptosis.[144]

There is a lot of evidence that hyperglycemia-induced alterations in the redox state are a key event in the development of diabetic retinopathy. The reason is that the retina has a high content of polyunsaturated fatty acids and the highest oxygen uptake and glucose oxidation of any other tissue, and these particularities make it especially susceptible to OS.[145] As a result, increased membrane lipid peroxidation and oxidation-induced damage to DNA have been described in the diabetic retina.[146,147]

Recently, the contribution of OS to the development of retinopathy has been reviewed comprehensively.[131] According to Kowluru and coworkers,[146,148,149] retinal OS is increased because of overproduction of ROS[2] in conjunction with decreased antioxidant defense mechanisms such as diminished activities of superoxide dismutase, glutathione, glutathione reductase, glutathione peroxidase, and catalase. Aforementioned pathomechanisms that play an important role in the development of diabetic retinopathy are the polyol pathway, hexosamine pathway, activation of PKC, and increased formation of AGEs.

The hexosamine pathway influences the posttranslational modification of intracellular factors,[150] whereas the activation of PKC contributes to the development of diabetic retinopathy[132,151] by increasing vascular permeability, stimulating neovascularization, promoting apoptosis, and altering endothelial proliferation.[152–154]

Besides augmenting the 4 aforementioned pathomechanisms,[2] ROS also alter the expression of vascular endothelial growth factor (VEGF)[155] and insulin-like growth factor-1 (IGF-1).[156]

VEGF induces angiogenesis and plays a major and detrimental role in the development of nonproliferative and proliferative diabetic retinopathy.[157] ROS not only increase VEGF expression in the retina[158] but further mediate its pathological effects on microvascular cells. Further vascular-specific growth factors essential for vascular development and maturation have been incriminated in the pathogenesis of diabetic retinopathy: platelet-derived growth factor, angiopoietins,[159] TGF-β, and pigment epithelium-derived growth factor.[160]

There are several angiopoietins, termed 1 to 4, that bind to Tie receptors (tyrosine kinases containing immunoglobulin and epidermal growth factor domains). Angiopoietin 1 (Ang-1) is produced by perivascular cells, whereas angiopoietin 2 (Ang-2) is secreted by endothelial cells. Ang-2 antagonizes Ang-1, and the latter decreases Ang-2 production indirectly by decreasing FOXO (Forkhead) activity. Ang-1 is anti-inflammatory toward endothelial cells, suppressing vascular leakage, and is believed to promote their survival and stabilization.[159] The current hypothesis is that VEGF is an important initiator of angiogenesis, whereas Ang-1 is mandatory for vascular development and stabilization, has anti-inflammatory effects, and protects against the development of diabetic retinopathy. Ang-2 has a destabilizing role in vascular remodeling and in lymphatic vessel growth[161] and is implicated in the early pericyte loss.[138]

The altered interaction between leukocytes and endothelial cells promotes local inflammation. Increased leukocyte adhesion is mediated directly via VEGF and RAGE, which acts also as a ligand for β_2-integrins. These are adhesion receptors located on the surface of monocytes, neutrophils, and lymphocytes.[162] However, RAGE can also contribute indirectly to leukocyte adhesion: RAGE activation by circulating AGE increases the expression of adhesion molecules on the surface of endothelial cells, thus promoting recruitment of inflammatory cells.[163]

The role of IGF-1 in the pathogenesis of diabetic retinopathy is not clear so far, but IGF-1 has mitogenic effects on endothelial cells, stimulates their glucose uptake by a PKC-dependent mechanism, and can modulate the expression of VEGF.[156] IGF-1 production itself can be up-regulated by ROS and the activation of PKC.

Caspases are a group of cysteine proteases involved in apoptosis[164] that are highly activated by oxidative and nitrative stress.[165] Damage to the mitochondrial membrane, resulting in leakage of cytochrome C, leads to activation of a caspase-dependent cascade that will culminate with the activation of caspase-3, which is responsible for DNA fragmentation.[166]

Two further pathomechanisms initiated by ROS are the activation of NF-κB[19] and the activation of both Ras and Rac,[167,168] 2 proteins belonging to the small GTPase family that play an important role in the regulation of growth, senescence, and cell transformation.[168] The importance of activated Ras and NF-κB in the pathogenesis of diabetic retinopathy has been highlighted.[167,169] As previously mentioned, redox-sensitive NF-κB has a dual effect, promoting either apoptosis or cell survival according to the cell type and stage of disease. In diabetic retinal endothelial cells and pericytes,[170] NF-κB activation is responsible for hyperglycemia-induced apoptosis.[25,169] Activation of NF-κB also triggers the release of inflammatory mediators such as TNF-α and activates iNOS, thus promoting OS.[71] Indeed, inflammatory cytokines [interleukin 1β (IL-1β), IL-6, and IL-8] are increased in the vitreous of patients with proliferative retinopathy.[171] IL-1β increases apoptosis by activating NF-κB (continuous activating loop)[172] and caspase-3 activates cyclooxygenase-1, thus increasing prostaglandin E_2. The latter modulates VEGF, contributing to increased vascular permeability and angiogenesis.[173] Activated macrophages seem to play an important role in the development of diabetic retinopathy by increased production of matrix metalloproteinase. The latter contributes to both abnormal retinal revascularization and the early steps of pericyte loss.[159]

ROS can also profoundly alter mitochondrial function by damaging the lipid membrane or even mitochondrial DNA.[174] Lipid membrane modification results in permeability changes, swelling of mitochondria,[174] the escape of some mitochondrial proteins such as cytochrome C into the cytosole, and the penetration of cytosolic proteins (eg, Bax) into mitochondria. These changes are thought to promote cell apoptosis.[175]

Diabetic retinopathy is not only driven by changes in pericytes and retinal endothelial cells: other cell types (Muller cells, ganglion cells, astrocytes, and photoreceptors) are also affected by diabetes-induced changes.[176–178]

Recently, homocysteine has emerged as an important marker of cardiovascular disease[179] and has been linked to vaso-occlusive disease of the eye.[180] Higher plasma homocysteine levels were found in diabetic retinopathy, and an increase by 1μ mol/L in the plasma homocysteine concentration independently predicted an increased risk for diabetic retinopathy of 15% to 20%.[181] However, homocysteine is currently regarded as a biomarker rather than a cause, and the exact pathomechanisms that connect hyperhomocysteinemia to diabetic retinopathy are not known.

Diabetic Neuropathy

The prevalence of diabetic neuropathy increases with the duration of diabetes and varies from 10% (1 year after diagnosis) to 50% (25 years after diagnosis),[182–184] with an average prevalence of 30%.[185] Neuropathy accompanying type 1 diabetes has a faster progression and more severe outcomes.[186] The underlying causes are related to hyperglycemia, insulin, and C-peptide deficiency and imply increased activity of the polyol pathway, decreased endoneurial blood flow, impaired neuronal Na^+/K^+-ATPase activity, and decreased NO production.[187] This is accompanied by a progressive decline in neurotropic factors such as nerve growth factor and insulin-like growth factor.[184,187]

Microangiopathic changes are also typical in the development of diabetic neuropathy.[188] Endoneurial microangiopathy parallels fiber loss[189] and is characterized by basement membrane thickening and bloated endothelial cells.[188]

Impaired Na^+/K^+-ATPase activity results in increased intra-axonal accumulation of Na^+ and reversible axonal and Ranvier's node swelling.[190,191]

Structural changes at the level of nodal and paranodal apparatus have been described in experimental diabetic neuropathy with a lateralization of nodal Na^+ channels resulting in a conduction block, an effect prevented by treatment with an aldose reductase inhibitor.[192]

Increased clearance of glucose through the polyol pathway results in an accumulation of fructose within the neuronal cells, causing a decrease in the intracellular concentration of myoinositol and taurine. Myoinositol depletion has been related to impaired PKC function and decreased activation of Na^+/K^+-ATPase.[193] Reduced Na^+/K^+-ATPase activity in peripheral nerves is believed to play a central role in the early electrophysiological, metabolic, and morphological abnormalities of experimental diabetic neuropathy.[194] The β-amino acid taurine has several physiological functions, such as regulating cell volume as an osmolyte in the heart and in the retina and possibly scavenging intracellular carbonyl groups and ROS.[195]

Activation of the polyol pathway also depletes NADPH, the cofactor of aldose reductase, glutathione reductase, and eNOS. Interestingly, the degeneration of intraepidermal nerve fibers is more pronounced in patients with high aldose reductase levels than in those with lower levels,[196] and the same is true for the prevalence of diabetic complications.[197]

Glycation of proteins within nerves in patients with diabetes[198] might destabilize the cytoskeleton and contribute to slowing of axonal transport and accelerated nerve degeneration.[199] Apart from neurons, structural proteins in Schwann cells (laminin) are also modified by glycation,[200] which results in impaired neuronal regeneration.

There is some evidence that antioxidants such as α-lipoid acid may be helpful in treating patients with diabetic neuropathy.[7]

Diabetic Nephropathy

Diabetic nephropathy, the leading cause of the development of end-stage renal disease, is structurally characterized by alterations of the glomerular filtration barrier (glomerular endothelial dysfunction, basement membrane thickening, and podocyte loss) and alterations of mesangial cells and subsequent accumulation of matrix in the mesangial area.[201] The accumulation of matrix in the mesangium results in a reduction of the capillary area available for filtration and progressive loss of renal function.[202]

Both environmental and genetic factors contribute to the susceptibility of developing nephropathy and influence the rate of progression.[203,204]

Increased protein filtration resulting in proteinuria stems from alterations of the glomerular filtration barrier, and it can be postulated that morphological changes such as decreased fenestration of glomerular endothelial cells, basement membrane thickening, and podocyte rarefaction precede the clinical onset of nephropathy by years.[205]

Decreased fenestration of glomerular endothelium has been suggested to occur early in the development of diabetic nephropathy.[206,207] It has been recently described in patients with type 1 diabetes mellitus[208] and may share common pathomechanisms with podocyte lesions.

Thickening of the glomerular basement membrane (GBM) occurs early after the onset of diabetes (1–2 years) because of increased extracellular matrix protein synthesis and impaired protein degradation.[209] GBM thickening is partly due to an increase in collagen IV,[210] which actually leads to a more porous structure allowing protein leakage.[211] Another mechanism facilitating protein loss is the decrease in negatively charged proteoglycans within the GBM, leading to a loss of its charge selectivity.[212,213] However, these changes seem to occur rather late in the development of diabetic nephropathy.[214] Proteoglycan synthesis takes place in all 3 types of glomerular cells but predominantly in podocytes.[201] Perlecan and especially agrin are the 2 main proteoglycans within the GBM,[215] and the synthesis of the latter is altered by both hyperglycemia and angiotensin II (ANG-II)[215,216]; this partly explains the antiproteinuric effect of angiotensin converting enzyme (ACE) inhibitors.

It has been suggested that podocyte rarefaction is an early finding of diabetic nephropathy predicting the further development of proteinuria, the hallmark of diabetic nephropathy,[217] and it has been excellently reviewed in detail elsewhere.[201] Briefly, podocytes are anchored to the GBM by foot processes, and the space between the latter is filled by a porous membrane (the slit diaphragm) that is permeable to water and small solutes but largely impermeable to proteins under normal conditions.[218–220] Signaling coming from the healthy slit diaphragm prevents podocyte apoptosis.[221,222] One main component of the slit diaphragm is a transmembrane protein called nephrin,[223] which is also anchored to actin filaments of the podocyte cytoskeleton.[224,225] Nephrin seems to play a major role in maintaining podocyte integrity.[201] A reduction in nephrin occurs in nephrotic patients with type 1 diabetes,[226] and nephrinuria increases with increased glomerular damage.[227] Decreased nephrin levels are probably a result of enhanced urinary loss due to the synthesis of variant isoforms that lack a transmembrane domain.[228]

Beyond podocytopenia,[217,229–231] diabetic nephropathy is also characterized by podocyte foot widening.[232] The latter finding is due to an attempt of podocytes to cover the GBM partly denuded by the slit diaphragm and to reduce the filtration of water.[217,233]

Podocyte loss promotes the development of glomerulosclerosis by the formation of synechiae between the denuded GBM and Bowman's capsule[224] and results from increased apoptosis and cell detachment.[201] In fact, progressive podocyte foot process detachment has been recently described in patients with type 1 diabetes mellitus.[208]

Podocyte apoptosis seems to be largely driven by ANG-II, and the effect is mediated by TGF-β.[234] Treatment with ACE inhibitors can reduce the number of urinary podocytes.[235] Detachment of podocytes may be triggered by the down-regulation of the $\alpha 3\beta 1$-integrin receptor, the main adhesion complex of podocytes to the GBM.[218,219,236,237] The later complex is decreased in diabetes,[238–240] and this mechanism is probably due to hyperglyemia[238,240] and mediated by TGF-β1.[241,242]

Another mechanism for ANG-II–mediated glomerular damage appears to be the inhibition of nephrin expression, but the exact mechanisms of damage are not known.[201] Injured podocytes express AT1 and possibly AT2 receptors and thus mediate damage triggered by ANG-II.[224,243] On the other hand, hyperglycemia induces ANG-II formation within podocytes by up-regulating angiotensinogen expression.[224] Proteinuria also promotes ANG-II formation by increasing protein passage through the filtration barrier[244] and by mechanical stretching.[245] The latter mechanism is independent of ACE, the conversion from ANG-I to ANG-II being alternatively driven by chymase.[246]

Diabetes is accompanied by increased OS,[5] a state that increases angiotensinogen expression[245] and induces podocyte dysfunction.[247,248]

Podocytes and mesangial cells play a major role in the development of diabetic nephropathy, but enhanced synthesis of several growth factors and cytokines is also important for the pathogenesis. Within this category, IGF-1, monocyte chemoattractant protein-1, IL-6, and especially VEGF and TGF-β1 are of major importance.

VEGF is a factor for survival and angiogenesis and also contributes to the development of diabetic nephropathy. It increases the permeability of the glomerular filtration barrier to proteins and stimulates NOS, promoting vasodilatation with subsequent hyperfiltration.[249,250] One major source for VEGF production is podocytes,[201] especially in the presence of diabetes.[251,252] VEGF expression is enhanced by hyperglycemia,[253] TGF-β1,[253,254] ANG-II,[255,256] and hypoxia-inducible factor-1.[257]

Furthermore, AGEs might play an important role in the up-regulation of VEGF production. RAGE is overexpressed in podocytes of *db/db* mice,[252] and RAGE activation results in increased VEGF and

TGF-β1 expression, whereas the inhibition of the AGE-RAGE interaction by soluble RAGE markedly decreases VEGF production and ameliorates clinical parameters of glomerular function.[252] In a recent study, Susztak and coworkers[258] demonstrated that the scavenger receptor CD36 was increased in proximal tubular cells in human diabetic nephropathy and mediated tubular cell apoptosis induced by glycated and free fatty acids.[258]

Interestingly, podocytes not only seem to be a major source of VEGF but also are sensitive to VEGF action.[254,259] This autocrine loop is believed to have a cytoprotective function via decreasing intracellular calcium[259] and a not entirely understood mechanism that implies protection mediated by nephrin.[222] The blockade of endogenous VEGF synthesis results in a reversal of cytoprotection.[222,259]

On the other hand, VEGF stimulates podocytes to produce the α3-chain of collagen IV,[254] and this mechanism might be partly incriminated in GBM thickening in diabetes. The blockade of endogenous VEGF action by, for example, a specific inhibitor of VEGF receptor kinase (SU5416) reduces the expression of α3(IV) collagen, preserves nephrin, and prevents albuminuria in diabetic *db/db* mice.[201]

TGF-β1 is a cytokine that is overexpressed in the kidneys in diabetes and has hypertrophic and sclerotic properties,[260] causing mesangial expansion, interstitial fibrosis, renal insufficiency, and proteinuria.[261] Still, its exact role in the development of diabetic nephropathy is unclear because its blockade does not significantly reverse albuminuria, and this suggests that inhibition of further factors (eg, VEGF) might be necessary.[262] Another hypothesis states that albuminuria is not necessarily related to mesangial matrix expansion and is probably more closely related to changes induced by VEGF and protected by nephrin, whereas TGF-β is implicated in the pathogenesis of glomerulosclerosis.[201]

On podocytes, TGF-β1 causes detachment and/or apoptosis,[234,238,263] stimulates the production of α3(IV) collagen but not α1(IV) and α5(IV) collagen, and increases VEGF synthesis.[253] TGF-β production by podocytes does not appear to be enhanced in diabetes, whereas TGF-β type II receptor expression is increased, and this suggests that these cells are more susceptible to TGF-β action in diabetes.[253]

Angiopoietins are involved in the formation and maturation of new blood vessels. Ang-1 is an angiogenesis-promoting factor sustaining the attachment of mesenchymal cells to endothelial tubes and the differentiation to mature pericytes. It thereby facilitates the creation of so-called nonleaky blood vessels.[264] Ang-2 antagonizes Ang-1, loosens the attachment of pericytes, and by synergizing with VEGF, promotes sprouting angiogenesis. Ang-2 is up-regulated in diabetic nephropathy[142] and along with VEGF may contribute to increased angiogenesis, reduced vessel maturation, and increased leakage.[79]

Inflammatory mechanisms mediated by macro phages and angiogenesis may further stimulate the development of diabetic nephropathy. Increased levels of VEGF and angiopoietin not only determine vessel leakage and reduce transendothelial electric resistance but also correlate with increased neovascularization.[218,265,266] VEGF induces activation of metalloproteinase and activation and migration of endothelial cells.[267]

The differential expression of cytokines is subject to modulation not only by diabetes but also by the patient's genetic background. Indeed, there is growing evidence for the implication of genetic factors in the pathogenesis of diabetic nephropathy.[268] For example, a decrease in VEGF and nephrin was found at the level of proteins and RNA, but the expression of genes encoding them has also been found to be decreased.[269] Because the analysis of human kidney tissue is extremely difficult for ethical reasons, most of the data available have an animal provenance.[205] It has been shown that peroxisome proliferator-activated receptor-α is up-regulated in glomeruli, cortical tubules, and renal arterial vessels of *db/db* mice, also triggering the up-regulation of other proteins (eg, adipocyte differentiation-regulated protein).[270] Another study showed that in nonobese diabetic mice changed gene expression occurred especially in structures implicated in 4 major signaling pathways: insulin, TGF-β, TNF-α, and peroxisome proliferator-activated receptor.[271]

CONCLUDING REMARKS

Over the past decades, strategies aiming at reducing diabetic complications have focused not only on optimizing the metabolic control and normalizing blood glucose levels but also on the reduction/blockade of hyperglycemia-induced pathomechanisms. Because many interventions in humans have been a disappointment, even though they have produced very promising results in animal studies, new strategies have been developed to improve the benefits and the clinical outcome of diabetic patients.

Inhibition of aldose reductase and therefore of the polyol pathway seemed, from a biochemical point of view, an appropriate approach,[272] and data from animal studies generated much hope. However, clinical trials of aldose reductase inhibitors were disappointing, and several pharmaceutical companies have abandoned the development of this line of drugs.[273]

The blockade of the hexosamine pathway is promising because increased activity of glutamine fructose-6 phosphate amidotransferase has been described in diabetes and associated with insulin resistance, postprandial hyperglycemia, and OS.[274] Clinical data are pending.

Inhibition of PKC-β by ruboxistaurin, a PKC-β isoform–selective inhibitor, normalizes endothelial dysfunction and the renal glomerular filtration rate and prevents visual acuity loss in people with diabetes. Therefore, PKC isoform–selective inhibitors are an interesting new class of agents, carrying the promise of delaying or preventing diabetic vascular disease.[129]

Another major pathomechanism, AGE generation, has gained importance over the past decades. Although the role of AGEs in the development of diabetes complications has become increasingly clear, we are far from finding a specific therapy. Because food AGEs are major contributors to the body AGE pool[275,276] and have detrimental in vivo effects,[277,278] reduction of the AGE burden might be another useful approach to diminishing the vascular risk in diabetes. AGE concentrations in food can be largely controlled by the selection of food and cooking methods.[279] Therefore, a more specific dietary approach might represent a simple and inexpensive method of influencing the development of diabetes complications. Still, data from interventional studies aiming at reducing AGE load need to be generated.

If the unifying hypothesis is valid, a more general approach is afforded not only to inhibit a single specific pathogenetic mechanisms but also to block the initial cause that leads to activation of the 4 main mechanisms discussed in the introduction that are responsible for establishing the complex network of processes involved in the development of macrovascular and microvascular complications. In our mind, treatment with benfotiamine bears some hope[133,280] because it is able to prevent the accumulation of C3 and C6 intermediates of glycolysis and therefore the activation of the main pathomechanisms driving the development of complications in diabetes. Unfortunately, long-term clinical data of its influence on AGE production and the prevention of complications are not yet available.

Reduction of OS may represent another target that should principally be able to reduce the vascular risk of patients with diabetes. This topic has been addressed in numerous studies, but the results have been disappointing, as already mentioned.[281] However, it might be that the used antioxidants, such as vitamin E and vitamin C, are too unspecific in their action profile and do not have access to the physiological site where ROS are generated or that the intervention is applied too late to prevent irreversible damage. We need a better understanding of why, where, and how ROS are generated in order to develop more specific approaches.

Lastly, the genetic background may play a major role in the development of diabetic complications because the rate of complication development is different even with comparable metabolic and risk factor control.[282] It might be that we have to differentiate between various subgroups of patients because the sensitivity and underlying mechanisms might vary with the genetic background. The elucidation of risk genes is advancing rapidly and may allow the development of specific, effective prophylactic strategies in the near future.

In conclusion, the mechanisms that are thought to generate diabetic complications present ample opportunity for further study and for potential targets in therapeutic trials.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. King H, Aubert RE, Herman WH. Global burden of diabetes, 1995–2025: prevalence, numerical estimates, and projections. *Diabetes Care* 2005; 1998: 1414–1431.
2. Brownlee M. The pathobiology of diabetic complications: a unifying mechanism. *Diabetes* 2005; 54: 1615–1625.
3. Kaiser N, Sasson S, Feener EP, et al. Differential regulation of glucose transport and transporters by glucose in vascular endothelial and smooth muscle cells. *Diabetes* 1993; 42: 80–89.
4. Gabbay KH, Merola LO, Field RA. Sorbitol pathway: presence in nerve and cord with substrate accumulation in diabetes. *Science* 1966; 151: 209–210.
5. Ceriello A, Motz E. Is oxidative stress the pathogenic mechanism underlying insulin resistance, diabetes, and cardiovascular disease? The common soil hypothesis revisited. *Arterioscler Thromb Vasc Biol* 2004; 24: 816–823.
6. Ceriello A. New insights on oxidative stress and diabetic complications may lead to a "causal" antioxidant therapy. *Diabetes Care* 2003; 26: 1589–1596.
7. Rosen P, Nawroth PP, King G, et al. The role of oxidative stress in the onset and progression of

diabetes and its complications: a summary of a congress series sponsored by UNESCO-MCBN, the American Diabetes Association and the German Diabetes Society. *Diabetes Metab Res Rev* 2001; 17: 189–212.

8. Cai W, Gao QD, Zhu L, et al. Oxidative stress-inducing carbonyl compounds from common foods: novel mediators of cellular dysfunction. *Mol Med* 2002; 8: 337–346.
9. Kimura S, Zhang GX, Nishiyama A, et al. Role of NAD(P)H oxidase-and mitochondria-derived reactive oxygen species in cardioprotection of ischemic reperfusion injury by angiotensin II. *Hypertension* 2005; 45: 860–866.
10. Zhang DX, Gutterman DD. Mitochondrial reactive oxygen species-mediated signaling in endothelial cells. *Am J Physiol Heart Circ Physiol* 2007; 292: H2023–H2031.
11. Brownlee M. Biochemistry and molecular cell biology of diabetic complications. *Nature* 2001; 414: 813–820.
12. Nishikawa T, Edelstein D, Brownlee M. The missing link: a single unifying mechanism for diabetic complications. *Kidney Int Suppl* 2000; 77: S26–S30.
13. Li JM, Shah AM. Endothelial cell superoxide generation: regulation and relevance for cardiovascular pathophysiology. *Am J Physiol Regul Integr Comp Physiol* 2004; 287: R1014–R1030.
14. Dandona P, Thusu K, Cook S, et al. Oxidative damage to DNA in diabetes mellitus. *Lancet* 1996; 347: 444–445.
15. Gopaul NK, Anggard EE, Mallet AI, et al. Plasma 8-epi-PGF2 alpha levels are elevated in individuals with non-insulin dependent diabetes mellitus. *FEBS Lett* 1995; 368: 225–229.
16. Sampson MJ, Gopaul N, Davies IR, et al. Plasma F2 isoprostanes: direct evidence of increased free radical damage during acute hyperglycemia in type 2 diabetes. *Diabetes Care* 2002; 25: 537–541.
17. Wiernsperger NF. Oxidative stress as a therapeutic target in diabetes: revisiting the controversy. *Diabetes Metab* 2003; 29: 579–585.
18. Nourooz-Zadeh J, Tajaddini-Sarmadi J, McCarthy S, et al. Elevated levels of authentic plasma hydroperoxides in NIDDM. *Diabetes* 1995; 44: 1054–1058.
19. Nourooz-Zadeh J, Rahimi A. Tajaddini-Sarmadi J, et al. Relationships between plasma measures of oxidative stress and metabolic control in NIDDM. *Diabetologia* 1997; 40: 647–653.
20. Dominguez C, Ruiz E, Gussinye M, Carrascosa A. Oxidative stress at onset and in early stages of type 1 diabetes in children and adolescents. *Diabetes Care* 1998; 21: 1736–1742.
21. Esposito K, Nappo F, Marfella R, et al. Inflammatory cytokine concentrations are acutely increased by hyperglycemia in humans: role of oxidative stress. *Circulation* 2002; 106: 2067–2072.
22. Facchini FS, Hua NW, Reaven GM, Stoohs RA. Hyperinsulinemia: the missing link among oxidative stress and age-related diseases? *Free Radic Biol Med* 2000; 29: 1302–1306.
23. Facchini FS, Humphreys MH, DoNascimento CA, et al. Relation between insulin resistance and plasma concentrations of lipid hydroperoxides, carotenoids, and tocopherols. *Am J Clin Nutr* 2000; 72: 776–779.
24. Mueller CF, Widder JD, McNally JS, et al. The role of the multidrug resistance protein-1 in modulation of endothelial cell oxidative stress. *Circ Res* 2005; 97: 637–644.
25. Du X, Stocklauser-Farber K, Rosen P. Generation of reactive oxygen intermediates, activation of NF-κB, and induction of apoptosis in human endothelial cells by glucose: role of nitric oxide synthase? *Free Radic Biol Med* 1999; 27: 752–763.
26. Pagano PJ, Ito Y, Tornheim K, et al. An NADPH oxidase superoxide-generating system in the rabbit aorta. *Am J Physiol* 1995; 268: H2274–H2280.
27. Droge W. Free radicals in the physiological control of cell function. *Physiol Rev* 2002; 82: 47–95.
28. Cadenas E, Davies KJ. Mitochondrial free radical generation, oxidative stress, and aging. *Free Radic Biol Med* 2000; 29: 222–230.
29. Finkel T. Radical medicine: treating ageing to cure disease. *Nat Rev Mol Cell Biol* 2005; 6: 971–976.
30. Raha S, Robinson BH. Mitochondria, oxygen free radicals, disease and ageing. *Trends Biochem Sci* 2000; 25: 502–508.
31. Trumpower BL. The protonmotive Q cycle. Energy transduction by coupling of proton translocation to electron transfer by the cytochrome bc1 complex. *J Biol Chem* 1990; 265: 11409–11412.
32. Wallace DC. Diseases of the mitochondrial DNA. *Annu Rev Biochem* 1992; 61: 1175–1212.
33. Du XL, Edelstein D, Dimmeler S, et al. Hyperglycemia inhibits endothelial nitric oxide synthase activity by posttranslational modification at the Akt site. *J Clin Invest* 2001; 108: 1341–1348.
34. Salonen JT, Nyyssonen K, Tuomainen TP, et al. Increased risk of non-insulin dependent diabetes mellitus at low plasma vitamin E concentrations: a four year follow up study in men. *BMJ* 1995; 311: 1124–1127.
35. Thornalley PJ, McLellan AC, Lo TW, et al. Negative association between erythrocyte reduced glutathione concentration and diabetic complications. *Clin Sci (Lond)* 1996; 91: 575–582.
36. Simon-Schnass I, Rosack C, Rösen P, et al. Risk of multiple micro-nutrient deficiency in the diet of type 2 diabetic subjects. *Curr Top Nutraceut Res* 2003; 1: 97–104.
37. Rosen P, Ballhausen T, Bloch W, Addicks K. Endothelial relaxation is disturbed by oxidative stress in the diabetic rat heart: influence of

tocopherol as antioxidant. *Diabetologia* 1995; 38: 1157–1168.

38. Kowluru RA, Kowluru V, Xiong Y, Ho YS. Overexpression of mitochondrial superoxide dismutase in mice protects the retina from diabetes-induced oxidative stress. *Free Radic Biol Med* 2006; 41: 1191–1196.
39. Lonn E, Yusuf S, Hoogwerf B, et al. Effects of vitamin E on cardiovascular and microvascular outcomes in high-risk patients with diabetes: results of the HOPE study and MICRO-HOPE substudy. *Diabetes Care* 2002; 25: 1919–1927.
40. Darko D, Dornhorst A, Kelly FJ, et al. Lack of effect of oral vitamin C on blood pressure, oxidative stress and endothelial function in type II diabetes. *Clin Sci (Lond)* 2002; 103: 339–344.
41. Bierhaus A, Humpert PM, Morcos M, et al. Understanding RAGE, the receptor for advanced glycation end products. *J Mol Med* 2005; 83: 876–886.
42. Barnes PJ, Karin M. Nuclear factor-kappaB: a pivotal transcription factor in chronic inflammatory diseases. *N Engl J Med* 1997; 336: 1066–1071.
43. Li J, Schmidt AM. Characterization and functional analysis of the promoter of RAGE, the receptor for advanced glycation end products. *J Biol Chem* 1997; 272: 16498–16506.
44. Soriano FG, Virag L, Szabo C. Diabetic endothelial dysfunction: role of reactive oxygen and nitrogen species production and poly(ADP-ribose) polymerase activation. *J Mol Med* 2001; 79: 437–448.
45. Soriano FG, Nogueira AC, Caldini EG, et al. Potential role of poly(adenosine 5′-diphosphate-ribose) polymerase activation in the pathogenesis of myocardial contractile dysfunction associated with human septic shock. *Crit Care Med* 2006; 34: 1073–1079.
46. Du XL, Edelstein D, Rossetti L, et al. Hyperglycemia-induced mitochondrial superoxide overproduction activates the hexosamine pathway and induces plasminogen activator inhibitor-1 expression by increasing Sp1 glycosylation. *Proc Natl Acad Sci U S A* 2000; 97: 12222–12226.
47. Wells L, Hart GW. O-GlcNAc turns twenty: functional implications for post-translational modification of nuclear and cytosolic proteins with a sugar. *FEBS Lett* 2003; 546: 154–158.
48. Lee AY, Chung SS. Contributions of polyol pathway to oxidative stress in diabetic cataract. *FASEB J* 1999; 13: 23–30.
49. Ceriello A, Morocutti A, Mercuri F, et al. Defective intracellular antioxidant enzyme production in type 1 diabetic patients with nephropathy. *Diabetes* 2000; 49: 2170–2177.
50. Gabbay KH. Hyperglycemia, polyol metabolism, and complications of diabetes mellitus. *Annu Rev Med* 1975; 26: 521–536.
51. Obrosova I, Faller A, Burgan J, et al. Glycolytic pathway, redox state of NAD(P)-couples and energy metabolism in lens in galactose-fed rats: effect of an aldose reductase inhibitor. *Curr Eye Res* 1997; 16: 34–43.
52. Idris I, Donnelly R. Protein kinase C beta inhibition: a novel therapeutic strategy for diabetic microangiopathy. *Diab Vasc Dis Res* 2006; 3: 172–178.
53. Wagner L, Laczy B, Tamasko M, et al. Cigarette smoke-induced alterations in endothelial nitric oxide synthase phosphorylation: role of protein kinase C. *Endothelium* 2007; 14: 245–255.
54. Pieper GM, Riaz UH. Activation of nuclear factor-kappaB in cultured endothelial cells by increased glucose concentration: prevention by calphostin C. *J Cardiovasc Pharmacol* 1997; 30: 528–532.
55. Giardino I, Edelstein D, Brownlee M. Nonenzymatic glycosylation in vitro and in bovine endothelial cells alters basic fibroblast growth factor activity. A model for intracellular glycosylation in diabetes. *J Clin Invest* 1994; 94: 110–117.
56. McLellan AC, Thornalley PJ, Benn J, Sonksen PH. Glyoxalase system in clinical diabetes mellitus and correlation with diabetic complications. *Clin Sci (Lond)* 1994; 87: 21–29.
57. Charonis AS, Reger LA, Dege JE, et al. Laminin alterations after in vitro nonenzymatic glycosylation. *Diabetes* 1990; 39: 807–814.
58. Vlassara H, Cai W, Crandall J, et al. Inflammatory mediators are induced by dietary glycotoxins, a major risk factor for diabetic angiopathy. *Proc Natl Acad Sci U S A* 2002; 99: 15596–15601.
59. Vlassara H. Advanced glycation end-products and atherosclerosis. *Ann Med* 1996; 28: 419–426.
60. Lu C, He JC, Cai W, et al. Advanced glycation endproduct (AGE) receptor 1 is a negative regulator of the inflammatory response to AGE in mesangial cells. *Proc Natl Acad Sci U S A* 2004; 101: 11767–11772.
61. Pugliese G, Pricci F, Iacobini C, et al. Accelerated diabetic glomerulopathy in galectin-3/AGE receptor 3 knockout mice. *FASEB J* 2001; 15: 2471–2479.
62. Li JJ, Dickson D, Hof PR, Vlassara H. Receptors for advanced glycosylation endproducts in human brain: role in brain homeostasis. *Mol Med* 1998; 4: 46–60.
63. Vlassara H. Recent progress in advanced glycation end products and diabetic complications. *Diabetes* 1997; 46(suppl 2): S19–S25.
64. Stitt AW, He C, Vlassara H. Characterization of the advanced glycation end-product receptor complex in human vascular endothelial cells. *Biochem Biophys Res Commun* 1999; 256: 549–556.
65. Reddy GK. Cross-linking in collagen by nonenzymatic glycation increases the matrix stiffness in rabbit Achilles tendon. *Exp Diabesity Res* 2004; 5: 143–153.
66. McFarlane S, Glenn JV, Lichanska AM, et al. Characterisation of the advanced glycation endproduct receptor complex in the retinal

pigment epithelium. *Br J Ophthalmol* 2005; 89: 107–112.

67. Bucciarelli LG, Wendt T, Rong L, et al. RAGE is a multiligand receptor of the immunoglobulin superfamily: implications for homeostasis and chronic disease. *Cell Mol Life Sci* 2002; 59: 1117–1128.
68. Ramasamy R, Vannucci SJ, Yan SS, et al. Advanced glycation end products and RAGE: a common thread in aging, diabetes, neurodegeneration, and inflammation. *Glycobiology* 2005; 15: 16R–28R.
69. Buetler TM, Leclerc E, Baumeyer A, et al. N(epsilon)-carboxymethyllysine-modified proteins are unable to bind to RAGE and activate an inflammatory response. *Mol Nutr Food Res* 2008; 52: 370–378.
70. Wever RM, Luscher TF, Cosentino F, Rabelink TJ. Atherosclerosis and the two faces of endothelial nitric oxide synthase. *Circulation* 1998; 97: 108–112.
71. Griscavage JM, Wilk S, Ignarro LJ. Inhibitors of the proteasome pathway interfere with induction of nitric oxide synthase in macrophages by blocking activation of transcription factor NF-κB. *Proc Natl Acad Sci U S A* 1996; 93: 3308–3312.
72. Mohamed AK, Bierhaus A, Schiekofer S, et al. The role of oxidative stress and NF-κB activation in late diabetic complications. *Biofactors* 1999; 10: 157–167.
73. Kowluru RA. Effect of advanced glycation end products on accelerated apoptosis of retinal capillary cells under in vitro conditions. *Life Sci* 2005; 76: 1051–1060.
74. Cipollone F, Iezzi A, Fazia M, et al. The receptor RAGE as a progression factor amplifying arachidonate-dependent inflammatory and proteolytic response in human atherosclerotic plaques: role of glycemic control. *Circulation* 2003; 108: 1070–1077.
75. Stern DM, Yan SD, Yan SF, Schmidt AM. Receptor for advanced glycation endproducts (RAGE) and the complications of diabetes. *Ageing Res Rev* 2002; 1: 1–15.
76. He CJ, Koschinsky T, Buenting C, Vlassara H. Presence of diabetic complications in type 1 diabetic patients correlates with low expression of mononuclear cell AGE-receptor-1 and elevated serum AGE. *Mol Med* 2001; 7: 159–168.
77. Cai W, He JC, Zhu L, et al. Advanced glycation end product (AGE) receptor 1 suppresses cell oxidant stress and activation signaling via EGF receptor. *Proc Natl Acad Sci U S A* 2006; 103: 13801–13806.
78. Vlassara H, Uribarri J. Glycoxidation and diabetic complications: modern lessons and a warning? *Rev Endocr Metab Disord* 2004; 5: 181–188.
79. Ichinose K, Kawasaki E, Eguchi K. Recent advancement of understanding pathogenesis of type 1 diabetes and potential relevance to diabetic nephropathy. *Am J Nephrol* 2007; 27: 554–564.
80. Strong JP, Mcgill HC Jr. The natural history of coronary atherosclerosis. *Am J Pathol* 1962; 40: 37–49.
81. Di MD, Mohn A, Mokini ZH, et al. Macroangiopathy in adults and children with diabetes: from molecular mechanisms to vascular damage (part 1). *Horm Metab Res* 2006; 38: 691–705.
82. Ross R. The pathogenesis of atherosclerosis: a perspective for the 1990s. *Nature* 1993; 362: 801–809.
83. Calles-Escandon J, Cipolla M. Diabetes and endothelial dysfunction: a clinical perspective. *Endocr Rev* 2001; 22: 36–52.
84. Holowatz LA, Thompson CS, Minson CT, Kenney WL. Mechanisms of acetylcholine-mediated vasodilatation in young and aged human skin. *J Physiol* 2005; 563: 965–973.
85. Garcia-Cardena G, Comander J, Anderson KR, et al. Biomechanical activation of vascular endothelium as a determinant of its functional phenotype. *Proc Natl Acad Sci U S A* 2001; 98: 4478–4485.
86. Dai G, Kaazempur-Mofrad MR, Natarajan S, et al. Distinct endothelial phenotypes evoked by arterial waveforms derived from atherosclerosis-susceptible and -resistant regions of human vasculature. *Proc Natl Acad Sci U S A* 2004; 101: 14871–14876.
87. Cunningham KS, Gotlieb AI. The role of shear stress in the pathogenesis of atherosclerosis. *Lab Invest* 2005; 85: 9–23.
88. Henrion D. Pressure and flow-dependent tone in resistance arteries. Role of myogenic tone. *Arch Mal Coeur Vaiss* 2005; 98: 913–921.
89. Shao JS, Cai J, Towler DA. Molecular mechanisms of vascular calcification: lessons learned from the aorta. *Arterioscler Thromb Vasc Biol* 2006; 26: 1423–1430.
90. Perrin RM, Harper SJ, Bates DO. A role for the endothelial glycocalyx in regulating microvascular permeability in diabetes mellitus. *Cell Biochem Biophys* 2007; 49: 65–72.
91. Tomkin GH. Targets for intervention in dyslipidemia in diabetes. *Diabetes Care* 2008; 31(suppl 2): S241–S248.
92. Bowie A, Owens D, Collins P, et al. Glycosylated low density lipoprotein is more sensitive to oxidation: implications for the diabetic patient? *Atherosclerosis* 1993; 102: 63–67.
93. Dimitriadis E, Griffin M, Owens D, et al. Oxidation of low-density lipoprotein in NIDDM: its relationship to fatty acid composition. *Diabetologia* 1995; 38: 1300–1306.
94. Deegan P, Owens D, Collins P, et al. Association between low-density lipoprotein composition and its metabolism in non-insulin-dependent diabetes mellitus. *Metabolism* 1999; 48: 118–124.
95. Gordon T, Castelli WP, Hjortland MC, et al. High density lipoprotein as a protective factor against coronary heart disease. The Framingham Study. *Am J Med* 1977; 62: 707–714.
96. Yu R, Yekta B, Vakili L, et al. Proatherogenic high-density lipoprotein, vascular inflammation, and

mimetic peptides. *Curr Atheroscler Rep* 2008; 10: 171–176.

97. Gomaraschi M, Calabresi L, Rossoni G, et al. Anti-inflammatory and cardioprotective activities of synthetic high-density lipoprotein containing apolipoprotein A-I mimetic peptides. *J Pharmacol Exp Ther* 2008; 324: 776–783.
98. Mackness MI, Durrington PN, Mackness B. The role of paraoxonase 1 activity in cardiovascular disease: potential for therapeutic intervention. *Am J Cardiovasc Drugs* 2004; 4: 211–217.
99. Hofer SE, Bennetts B, Chan AK, et al. Association between PON 1 polymorphisms, PON activity and diabetes complications. *J Diabet Complications* 2006; 20: 322–328.
100. Hu Y, Tian H, Liu R. Gln-Arg192 polymorphism of paraoxonase 1 is associated with carotid intima-media thickness in patients of type 2 diabetes mellitus of Chinese. *Diabetes Res Clin Pract* 2003; 61: 21–27.
101. Letellier C, Durou MR, Jouanolle AM, et al. Serum paraoxonase activity and paraoxonase gene polymorphism in type 2 diabetic patients with or without vascular complications. *Diabetes Metab* 2002; 28: 297–304.
102. Stasek JE Jr, Patterson CE, Garcia JG. Protein kinase C phosphorylates caldesmon 77 and vimentin and enhances albumin permeability across cultured bovine pulmonary artery endothelial cell monolayers. *J Cell Physiol* 1992; 153: 62–75.
103. Ramirez MM, Kim DD, Duran WN. Protein kinase C modulates microvascular permeability through nitric oxide synthase. *Am J Physiol* 1996; 271: H1702–H1705.
104. Abedin M, Tintut Y, Demer LL. Vascular calcification: mechanisms and clinical ramifications. *Arterioscler Thromb Vasc Biol* 2004; 24: 1161–1170.
105. Stary HC, Chandler AB, Dinsmore RE, et al. A definition of advanced types of atherosclerotic lesions and a histological classification of atherosclerosis. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. *Arterioscler Thromb Vasc Biol* 1995; 15: 1512–1531.
106. London GM, Marchais SJ, Guerin AP, Metivier F. Arteriosclerosis, vascular calcifications and cardiovascular disease in uremia. *Curr Opin Nephrol Hypertens* 2005; 14: 525–531.
107. Lehto S, Niskanen L, Suhonen M, et al. Medial artery calcification. A neglected harbinger of cardiovascular complications in non-insulin-dependent diabetes mellitus. *Arterioscler Thromb Vasc Biol* 1996; 16: 978–983.
108. Tanimura A, McGregor DH, Anderson HC. Matrix vesicles in atherosclerotic calcification. *Proc Soc Exp Biol Med* 1983; 172: 173–177.
109. Reynolds JL, Joannides AJ, Skepper JN, et al. Human vascular smooth muscle cells undergo vesicle-mediated calcification in response to changes in extracellular calcium and phosphate concentrations: a potential mechanism for accelerated vascular calcification in ESRD. *J Am Soc Nephrol* 2004; 15: 2857–2867.
110. Kim KM. Calcification of matrix vesicles in human aortic valve and aortic media. *Fed Proc* 1976; 35: 156–162.
111. Towler DA, Bidder M, Latifi T, et al. Diet-induced diabetes activates an osteogenic gene regulatory program in the aortas of low density lipoprotein receptor-deficient mice. *J Biol Chem* 1998; 273: 30427–30434.
112. Shao JS, Cheng SL, Pingsterhaus JM, et al. Msx2 promotes cardiovascular calcification by activating paracrine Wnt signals. *J Clin Invest* 2005; 115: 1210–1220.
113. Bujan J, Bellon JM, Sabater C, et al. Modifications induced by atherogenic diet in the capacity of the arterial wall in rats to respond to surgical insult. *Atherosclerosis* 1996; 122: 141–152.
114. Hu Y, Zhang Z, Torsney E, et al. Abundant progenitor cells in the adventitia contribute to atherosclerosis of vein grafts in ApoE-deficient mice. *J Clin Invest* 2004; 113: 1258–1265.
115. Zhang L, Zalewski A, Liu Y, et al. Diabetes-induced oxidative stress and low-grade inflammation in porcine coronary arteries. *Circulation* 2003; 108: 472–478.
116. Moreno PR, Fuster V. New aspects in the pathogenesis of diabetic atherothrombosis. *J Am Coll Cardiol* 2004; 44: 2293–2300.
117. Otto CM, Kuusisto J, Reichenbach DD, et al. Characterization of the early lesion of 'degenerative' valvular aortic stenosis. Histological and immunohistochemical studies. *Circulation* 1994; 90: 844–853.
118. McMahon R, Murphy M, Clarkson M, et al. IHG-2, a mesangial cell gene induced by high glucose, is human gremlin. Regulation by extracellular glucose concentration, cyclic mechanical strain, and transforming growth factor-beta1. *J Biol Chem* 2000; 275: 9901–9904.
119. Csiszar A, Smith KE, Koller A, et al. Regulation of bone morphogenetic protein-2 expression in endothelial cells: role of nuclear factor-kappaB activation by tumor necrosis factor-alpha, H2O2, and high intravascular pressure. *Circulation* 2005; 111: 2364–2372.
120. Moe SM, Chen NX. Pathophysiology of vascular calcification in chronic kidney disease. *Circ Res* 2004; 95: 560–567.
121. Collett GD, Canfield AE. Angiogenesis and pericytes in the initiation of ectopic calcification. *Circ Res* 2005; 96: 930–938.
122. Reynolds JL, Skepper JN, McNair R, et al. Multifunctional roles for serum protein fetuin-a in inhibition of human vascular smooth muscle cell calcification. *J Am Soc Nephrol* 2005; 16: 2920–2930.
123. Steitz SA, Speer MY, McKee MD, et al. Osteopontin inhibits mineral deposition and promotes regression of ectopic calcification. *Am J Pathol* 2002; 161: 2035–2046.
124. Proudfoot D, Davies JD, Skepper JN, et al. Acetylated low-density lipoprotein stimulates

human vascular smooth muscle cell calcification by promoting osteoblastic differentiation and inhibiting phagocytosis. *Circulation* 2002; 106: 3044–3050.

125. Bostrom K, Watson KE, Horn S, et al. Bone morphogenetic protein expression in human atherosclerotic lesions. *J Clin Invest* 1993; 91: 1800–1809.
126. Tyson KL, Reynolds JL, McNair R, et al. Osteo/chondrocytic transcription factors and their target genes exhibit distinct patterns of expression in human arterial calcification. *Arterioscler Thromb Vasc Biol* 2003; 23: 489–494.
127. Mokini Z, Chiarelli F. The molecular basis of diabetic microangiopathy. *Pediatr Endocrinol Rev* 2006; 4: 138–152.
128. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 1998; 352: 837–853.
129. Das EN, King GL. The role of protein kinase C activation and the vascular complications of diabetes. *Pharmacol Res* 2007; 55: 498–510.
130. Schalkwijk CG, Fa M, van HV, Stehouwer CD. Pathophysiological role of Amadori-glycated proteins in diabetic microangiopathy. *Semin Vasc Med* 2002; 2: 191–197.
131. Kowluru RA, Chan PS. Oxidative stress and diabetic retinopathy. *Exp Diabetes Res* 2007; 2007: 43603.
132. Koya D, King GL. Protein kinase C activation and the development of diabetic complications. *Diabetes* 1998; 47: 859–866.
133. Hammes HP, Du X, Edelstein D, et al. Benfotiamine blocks three major pathways of hyperglycemic damage and prevents experimental diabetic retinopathy. *Nat Med* 2003; 9: 294–299.
134. Asnaghi V, Gerhardinger C, Hoehn T, et al. A role for the polyol pathway in the early neuroretinal apoptosis and glial changes induced by diabetes in the rat. *Diabetes* 2003; 52: 506–511.
135. Heath H, Paterson RA, Hart JC. Changes in the hydroxyproline, hexosamine and sialic acid of the diabetic human and beta, beta'-iminodipropionitrile-treated rat retinal vascular systems. *Diabetologia* 1967; 3: 515–518.
136. Nakamura M, Barber AJ, Antonetti DA, et al. Excessive hexosamines block the neuroprotective effect of insulin and induce apoptosis in retinal neurons. *J Biol Chem* 2001; 276: 43748–43755.
137. Ola MS, Berkich DA, Xu Y, et al. Analysis of glucose metabolism in diabetic rat retinas. *Am J Physiol Endocrinol Metab* 2006; 290: E1057–E1067.
138. Hammes HP. Pericytes and the pathogenesis of diabetic retinopathy. *Horm Metab Res* 2005; 37(suppl 1): 39–43.
139. Robison WG Jr, Kador PF, Kinoshita JH. Early retinal microangiopathy: prevention with aldose reductase inhibitors. *Diabet Med* 1985; 2: 196–199.
140. Stitt AW, Li YM, Gardiner TA, et al. Advanced glycation end products (AGEs) co-localize with AGE receptors in the retinal vasculature of diabetic and of AGE-infused rats. *Am J Pathol* 1997; 150: 523–531.
141. Mizutani M, Kern TS, Lorenzi M. Accelerated death of retinal microvascular cells in human and experimental diabetic retinopathy. *J Clin Invest* 1996; 97: 2883–2890.
142. Hammes HP, Lin J, Wagner P, et al. Angiopoietin-2 causes pericyte dropout in the normal retina: evidence for involvement in diabetic retinopathy. *Diabetes* 2004; 53: 1104–1110.
143. Betsholtz C. Insight into the physiological functions of PDGF through genetic studies in mice. *Cytokine Growth Factor Rev* 2004; 15: 215–228.
144. Hammes HP, Lin J, Renner O, et al. Pericytes and the pathogenesis of diabetic retinopathy. *Diabetes* 2002; 51: 3107–3112.
145. Anderson RE, Rapp LM, Wiegand RD. Lipid peroxidation and retinal degeneration. *Curr Eye Res* 1984; 3: 223–227.
146. Kowluru RA, Tang J, Kern TS. Abnormalities of retinal metabolism in diabetes and experimental galactosemia. VII. Effect of long-term administration of antioxidants on the development of retinopathy. *Diabetes* 2001; 50: 1938–1942.
147. Kowluru RA. Diabetes-induced elevations in retinal oxidative stress, protein kinase C and nitric oxide are interrelated. *Acta Diabetol* 2001; 38: 179–185.
148. Haskins K, Bradley B, Powers K, et al. Oxidative stress in type 1 diabetes. *Ann N Y Acad Sci* 2003; 1005: 43–54.
149. Kern TS, Kowluru RA, Engerman RL. Abnormalities of retinal metabolism in diabetes or galactosemia: ATPases and glutathione. *Invest Ophthalmol Vis Sci* 1994; 35: 2962–2967.
150. Du X, Matsumura T, Edelstein D, et al. Inhibition of GAPDH activity by poly(ADP-ribose) polymerase activates three major pathways of hyperglycemic damage in endothelial cells. *J Clin Invest* 2003; 112: 1049–1057.
151. Ishii H, Jirousek MR, Koya D, et al. Amelioration of vascular dysfunctions in diabetic rats by an oral PKC beta inhibitor. *Science* 1996; 272: 728–731.
152. Oikawa T, Shimamura M, Ashino H, et al. Inhibition of angiogenesis by staurosporine, a potent protein kinase inhibitor. *J Antibiot (Tokyo)* 1992; 45: 1155–1160.
153. Xia P, Aiello LP, Ishii H, et al. Characterization of vascular endothelial growth factor's effect on the activation of protein kinase C, its isoforms, and endothelial cell growth. *J Clin Invest* 1996; 98: 2018–2026.
154. Koya D, Jirousek MR, Lin YW, et al. Characterization of protein kinase C beta isoform activation on the gene expression of transforming growth factor-beta, extracellular matrix components, and prostanoids in the

glomeruli of diabetic rats. *J Clin Invest* 1997; 100: 115–126.

155. Caldwell RB, Bartoli M, Behzadian MA, et al. Vascular endothelial growth factor and diabetic retinopathy: role of oxidative stress. *Curr Drug Targets* 2005; 6: 511–524.
156. DeBosch BJ, Baur E, Deo BK, et al. Effects of insulin-like growth factor-1 on retinal endothelial cell glucose transport and proliferation. *J Neurochem* 2001; 77: 1157–1167.
157. Aiello LP, Wong JS. Role of vascular endothelial growth factor in diabetic vascular complications. *Kidney Int Suppl* 2000; 77: S113–S119.
158. Lu M, Kuroki M, Amano S, et al. Advanced glycation end products increase retinal vascular endothelial growth factor expression. *J Clin Invest* 1998; 101: 1219–1224.
159. Bloomgarden ZT. Diabetic retinopathy and neuropathy. *Diabetes Care* 2005; 28: 963–970.
160. Fong DS, Aiello LP, Ferris FL III, Klein R. Diabetic retinopathy. *Diabetes Care* 2004; 27: 2540–2553.
161. Thurston G, Rudge JS, Ioffe E, et al. Angiopoietin-1 protects the adult vasculature against plasma leakage. *Nat Med* 2000; 6: 460–463.
162. May AE, Kanse SM, Lund LR, et al. Urokinase receptor (CD87) regulates leukocyte recruitment via beta 2 integrins in vivo. *J Exp Med* 1998; 188: 1029–1037.
163. Vlassara H, Fuh H, Donnelly T, Cybulsky M. Advanced glycation endproducts promote adhesion molecule (VCAM-1, ICAM-1) expression and atheroma formation in normal rabbits. *Mol Med* 1995; 1: 447–456.
164. Alnemri ES. Mammalian cell death proteases: a family of highly conserved aspartate specific cysteine proteases. *J Cell Biochem* 1997; 64: 33–42.
165. Mohr S, Zech B, Lapetina EG, Brune B. Inhibition of caspase-3 by S-nitrosation and oxidation caused by nitric oxide. *Biochem Biophys Res Commun* 1997; 238: 387–391.
166. Phaneuf S, Leeuwenburgh C. Cytochrome C release from mitochondria in the aging heart: a possible mechanism for apoptosis with age. *Am J Physiol Regul Integr Comp Physiol* 2002; 282: R423–R430.
167. Kowluru RA, Kowluru A, Chakrabarti S, Khan Z. Potential contributory role of H-Ras, a small G-protein, in the development of retinopathy in diabetic rats. *Diabetes* 2004; 53: 775–783.
168. Finkel T. Intracellular redox regulation by the family of small GTPases. *Antioxid Redox Signal* 2006; 8: 1857–1863.
169. Kowluru RA, Koppolu P, Chakrabarti S, Chen S. Diabetes-induced activation of nuclear transcriptional factor in the retina, and its inhibition by antioxidants. *Free Radic Res* 2003; 37: 1169–1180.
170. Romeo G, Liu WH, Asnaghi V, et al. Activation of nuclear factor-kappaB induced by diabetes and high glucose regulates a proapoptotic program in retinal pericytes. *Diabetes* 2002; 51: 2241–2248.
171. Yuuki T, Kanda T, Kimura Y, et al. Inflammatory cytokines in vitreous fluid and serum of patients with diabetic vitreoretinopathy. *J Diabet Complications* 2001; 15: 257–259.
172. Chang CK, LoCicero J III. Overexpressed nuclear factor kappaB correlates with enhanced expression of interleukin-1beta and inducible nitric oxide synthase in aged murine lungs to endotoxic stress. *Ann Thorac Surg* 2004; 77: 1222–1227.
173. Wilkinson-Berka JL. Vasoactive factors and diabetic retinopathy: vascular endothelial growth factor, cycoloxygenase-2 and nitric oxide. *Curr Pharm Des* 2004; 10: 3331–3348.
174. Kanwar M, Chan PS, Kern TS, Kowluru RA. Oxidative damage in the retinal mitochondria of diabetic mice: possible protection by superoxide dismutase. *Invest Ophthalmol Vis Sci* 2007; 48: 3805–3811.
175. Kowluru RA, Abbas SN. Diabetes-induced mitochondrial dysfunction in the retina. *Invest Ophthalmol Vis Sci* 2003; 44: 5327–5334.
176. Barber AJ, Lieth E, Khin SA, et al. Neural apoptosis in the retina during experimental and human diabetes. Early onset and effect of insulin. *J Clin Invest* 1998; 102: 783–791.
177. Mizutani M, Gerhardinger C, Lorenzi M. Muller cell changes in human diabetic retinopathy. *Diabetes* 1998; 47: 445–449.
178. Phipps JA, Fletcher EL, Vingrys AJ. Paired-flash identification of rod and cone dysfunction in the diabetic rat. *Invest Ophthalmol Vis Sci* 2004; 45: 4592–4600.
179. Wierzbicki AS. Homocysteine and cardiovascular disease: a review of the evidence. *Diab Vasc Dis Res* 2007; 4: 143–150.
180. Coral K, Raman R, Rathi S, et al. Plasma homocysteine and total thiol content in patients with exudative age-related macular degeneration. *Eye* 2006; 20: 203–207.
181. Brazionis L, Rowley K Sr, Itsiopoulos C, et al. Homocysteine and diabetic retinopathy. *Diabetes Care* 2008; 31: 50–56.
182. Pirart J. Diabetes mellitus and its degenerative complications: a prospective study of 4,400 patients observed between 1947 and 1973 [in French]. *Diabetes Metab* 1977; 3: 97–107.
183. Vinik AI, Holland MT, Le Beau JM, et al. Diabetic neuropathies. *Diabetes Care* 1992; 15: 1926–1975.
184. Sima AA, Zhang W, Grunberger G. Type 1 diabetic neuropathy and C-peptide. *Exp Diabesity Res* 2004; 5: 65–77.
185. Tesfaye S, Stevens LK, Stephenson JM, et al. Prevalence of diabetic peripheral neuropathy and its relation to glycaemic control and potential risk factors: the EURODIAB IDDM Complications Study. *Diabetologia* 1996; 39: 1377–1384.
186. Dyck PJ, Davies JL, Wilson DM, et al. Risk factors for severity of diabetic polyneuropathy: intensive

longitudinal assessment of the Rochester Diabetic Neuropathy Study cohort. *Diabetes Care* 1999; 22: 1479–1486.

187. Sima AA, Sugimoto K. Experimental diabetic neuropathy: an update. *Diabetologia* 1999; 42: 773–788.
188. Yagihashi S, Yamagishi S, Wada R. Pathology and pathogenetic mechanisms of diabetic neuropathy: correlation with clinical signs and symptoms. *Diabetes Res Clin Pract* 2007; 77(suppl 1): S184–S189.
189. Dyck PJ, Giannini C. Pathologic alterations in the diabetic neuropathies of humans: a review. *J Neuropathol Exp Neurol* 1996; 55: 1181–1193.
190. Brismar T, Sima AA. Changes in nodal function in nerve fibres of the spontaneously diabetic BB-Wistar rat: potential clamp analysis. *Acta Physiol Scand* 1981; 113: 499–506.
191. Sima AA, Brismar T. Reversible diabetic nerve dysfunction: structural correlates to electrophysiological abnormalities. *Ann Neurol* 1985; 18: 21–29.
192. Cherian PV, Kamijo M, Angelides KJ, Sima AA. Nodal Na(+)-channel displacement is associated with nerve-conduction slowing in the chronically diabetic BB/W rat: prevention by aldose reductase inhibition. *J Diabet Complications* 1996; 10: 192–200.
193. Ishii H, Koya D, King GL. Protein kinase C activation and its role in the development of vascular complications in diabetes mellitus. *J Mol Med* 1998; 76: 21–31.
194. Lattimer SA, Sima AA, Greene DA. In vitro correction of impaired Na+-K+-ATPase in diabetic nerve by protein kinase C agonists. *Am J Physiol* 1989; 256: E264–E269.
195. Hansen SH. The role of taurine in diabetes and the development of diabetic complications. *Diabetes Metab Res Rev* 2001; 17: 330–346.
196. Hirai A, Yasuda H, Joko M, et al. Evaluation of diabetic neuropathy through the quantitation of cutaneous nerves. *J Neurol Sci* 2000; 172: 55–62.
197. Nishimura C, Saito T, Ito T, et al. High levels of erythrocyte aldose reductase and diabetic retinopathy in NIDDM patients. *Diabetologia* 1994; 37: 328–330.
198. Vlassara H, Brownlee M, Cerami A. Excessive nonenzymatic glycosylation of peripheral and central nervous system myelin components in diabetic rats. *Diabetes* 1983; 32: 670–674.
199. Ryle C, Leow CK, Donaghy M. Nonenzymatic glycation of peripheral and central nervous system proteins in experimental diabetes mellitus. *Muscle Nerve* 1997; 20: 577–584.
200. Federoff HJ, Lawrence D, Brownlee M. Nonenzymatic glycosylation of laminin and the laminin peptide CIKVAVS inhibits neurite outgrowth. *Diabetes* 1993; 42: 509–513.
201. Wolf G, Chen S, Ziyadeh FN. From the periphery of the glomerular capillary wall toward the center of disease: podocyte injury comes of age in diabetic nephropathy. *Diabetes* 2005; 54: 1626–1634.
202. Mauer SM, Steffes MW, Ellis EN, et al. Structural-functional relationships in diabetic nephropathy. *J Clin Invest* 1984; 74: 1143–1155.
203. Ng DP, Krolewski AS. Molecular genetic approaches for studying the etiology of diabetic nephropathy. *Curr Mol Med* 2005; 5: 509–525.
204. Jones CA, Krolewski AS, Rogus J, et al. Epidemic of end-stage renal disease in people with diabetes in the United States population: do we know the cause? *Kidney Int* 2005; 67: 1684–1691.
205. Susztak K, Bottinger EP. Diabetic nephropathy: a frontier for personalized medicine. *J Am Soc Nephrol* 2006; 17: 361–367.
206. Deen WM. What determines glomerular capillary permeability? *J Clin Invest* 2004; 114: 1412–1414.
207. Deen WM, Lazzara MJ, Myers BD. Structural determinants of glomerular permeability. *Am J Physiol Renal Physiol* 2001; 281: F579–F596.
208. Toyoda M, Najafian B, Kim Y, et al. Podocyte detachment and reduced glomerular capillary endothelial fenestration in human type 1 diabetic nephropathy. *Diabetes* 2007; 56: 2155–2160.
209. Wolf G, Ziyadeh FN. Cellular and molecular mechanisms of proteinuria in diabetic nephropathy. *Nephron Physiol* 2007; 106: 26–31.
210. Kefalides NA. Basement membrane research in diabetes mellitus. *Coll Relat Res* 1981; 1: 295–299.
211. Isogai S, Mogami K, Shiina N, Yoshino G. Initial ultrastructural changes in pore size and anionic sites of the glomerular basement membrane in streptozotocin-induced diabetic rats and their prevention by insulin treatment. *Nephron* 1999; 83: 53–58.
212. Adler S. Structure-function relationships associated with extracellular matrix alterations in diabetic glomerulopathy. *J Am Soc Nephrol* 1994; 5: 1165–1172.
213. Kanwar YS, Liu ZZ, Kashihara N, Wallner EI. Current status of the structural and functional basis of glomerular filtration and proteinuria. *Semin Nephrol* 1991; 11: 390–413.
214. Vernier RL, Steffes MW, Sisson-Ross S, Mauer SM. Heparan sulfate proteoglycan in the glomerular basement membrane in type 1 diabetes mellitus. *Kidney Int* 1992; 41: 1070–1080.
215. Brinkkoetter PT, Holtgrefe S, van der Woude FJ, Yard BA. Angiotensin II type 1-receptor mediated changes in heparan sulfate proteoglycans in human SV40 transformed podocytes. *J Am Soc Nephrol* 2004; 15: 33–40.
216. Yard BA, Kahlert S, Engelleiter R, et al. Decreased glomerular expression of agrin in diabetic nephropathy and podocytes, cultured in

high glucose medium. *Exp Nephrol* 2001; 9: 214–222.

217. Pagtalunan ME, Miller PL, Jumping-Eagle S, et al. Podocyte loss and progressive glomerular injury in type II diabetes. *J Clin Invest* 1997; 99: 342–348.
218. Tsilibary EC. Microvascular basement membranes in diabetes mellitus. *J Pathol* 2003; 200: 537–546.
219. Kretzler M. Regulation of adhesive interaction between podocytes and glomerular basement membrane. *Microsc Res Tech* 2002; 57: 247–253.
220. Kojima K, Kerjaschki D. Is podocyte shape controlled by the dystroglycan complex? *Nephrol Dial Transplant* 2002; 17(suppl 9): 23–24.
221. Huber TB, Hartleben B, Kim J, et al. Nephrin and CD2AP associate with phosphoinositide 3-OH kinase and stimulate AKT-dependent signaling. *Mol Cell Biol* 2003; 23: 4917–4928.
222. Foster RR, Saleem MA, Mathieson PW, et al. Vascular endothelial growth factor and nephrin interact and reduce apoptosis in human podocytes. *Am J Physiol Renal Physiol* 2005; 288: F48–F57.
223. Khoshnoodi J, Sigmundsson K, Ofverstedt LG, et al. Nephrin promotes cell-cell adhesion through homophilic interactions. *Am J Pathol* 2003; 163: 2337–2346.
224. Pavenstadt H, Kriz W, Kretzler M. Cell biology of the glomerular podocyte. *Physiol Rev* 2003; 83: 253–307.
225. Saleem MA, Ni L, Witherden I, et al. Colocalization of nephrin, podocin, and the actin cytoskeleton: evidence for a role in podocyte foot process formation. *Am J Pathol* 2002; 161: 1459–1466.
226. Doublier S, Salvidio G, Lupia E, et al. Nephrin expression is reduced in human diabetic nephropathy: evidence for a distinct role for glycated albumin and angiotensin II. *Diabetes* 2003; 52: 1023–1030.
227. Patari A, Forsblom C, Havana M, et al. Nephrinuria in diabetic nephropathy of type 1 diabetes. *Diabetes* 2003; 52: 2969–2974.
228. Aaltonen P, Luimula P, Astrom E, et al. Changes in the expression of nephrin gene and protein in experimental diabetic nephropathy. *Lab Invest* 2001; 81: 1185–1190.
229. Steffes MW, Schmidt D, McCrery R, Basgen JM. Glomerular cell number in normal subjects and in type 1 diabetic patients. *Kidney Int* 2001; 59: 2104–2113.
230. White KE, Bilous RW, Marshall SM, et al. Podocyte number in normotensive type 1 diabetic patients with albuminuria. *Diabetes* 2002; 51: 3083–3089.
231. Dalla VM, Masiero A, Roiter AM, et al. Is podocyte injury relevant in diabetic nephropathy? Studies in patients with type 2 diabetes. *Diabetes* 2003; 52: 1031–1035.
232. Ellis EN, Steffes MW, Chavers B, Mauer SM. Observations of glomerular epithelial cell structure in patients with type I diabetes mellitus. *Kidney Int* 1987; 32: 736–741.
233. Nelson RG, Meyer TW, Myers BD, Bennett PH. Clinical and pathological course of renal disease in non-insulin-dependent diabetes mellitus: the Pima Indian experience. *Semin Nephrol* 1997; 17: 124–131.
234. Ding G, Reddy K, Kapasi AA, et al. Angiotensin II induces apoptosis in rat glomerular epithelial cells. *Am J Physiol Renal Physiol* 2002; 283: F173–F180.
235. Nakamura T, Ushiyama C, Suzuki S, et al. Urinary excretion of podocytes in patients with diabetic nephropathy. *Nephrol Dial Transplant* 2000; 15: 1379–1383.
236. Korhonen M, Ylanne J, Laitinen L, Virtanen I. Distribution of beta 1 and beta 3 integrins in human fetal and adult kidney. *Lab Invest* 1990; 62: 616–625.
237. Adler S. Characterization of glomerular epithelial cell matrix receptors. *Am J Pathol* 1992; 141: 571–578.
238. Chen HC, Chen CA, Guh JY, et al. Altering expression of alpha3beta1 integrin on podocytes of human and rats with diabetes. *Life Sci* 2000; 67: 2345–2353.
239. Regoli M, Bendayan M. Alterations in the expression of the alpha 3 beta 1 integrin in certain membrane domains of the glomerular epithelial cells (podocytes) in diabetes mellitus. *Diabetologia* 1997; 40: 15–22.
240. Kitsiou PV, Tzinia AK, Stetler-Stevenson WG, et al. Glucose-induced changes in integrins and matrix-related functions in cultured human glomerular epithelial cells. *Am J Physiol Renal Physiol* 2003; 284: F671–F679.
241. Kagami S, Border WA, Ruoslahti E, Noble NA. Coordinated expression of beta 1 integrins and transforming growth factor-beta-induced matrix proteins in glomerulonephritis. *Lab Invest* 1993; 69: 68–76.
242. Kumar NM, Sigurdson SL, Sheppard D, Lwebuga-Mukasa JS. Differential modulation of integrin receptors and extracellular matrix laminin by transforming growth factor-beta 1 in rat alveolar epithelial cells. *Exp Cell Res* 1995; 221: 385–394.
243. Gloy J, Henger A, Fischer KG, et al. Angiotensin II depolarizes podocytes in the intact glomerulus of the rat. *J Clin Invest* 1997; 99: 2772–2781.
244. Abbate M, Zoja C, Morigi M, et al. Transforming growth factor-beta1 is up-regulated by podocytes in response to excess intraglomerular passage of proteins: a central pathway in progressive glomerulosclerosis. *Am J Pathol* 2002; 161: 2179–2193.
245. Durvasula RV, Petermann AT, Hiromura K, et al. Activation of a local tissue angiotensin system in podocytes by mechanical strain. *Kidney Int* 2004; 65: 30–39.
246. Huang XR, Chen WY, Truong LD, Lan HY. Chymase is upregulated in diabetic nephropathy: implications for an alternative pathway of angiotensin II-mediated diabetic renal and

vascular disease. *J Am Soc Nephrol* 2003; 14: 1738–1747.

247. Vega-Warner V, Ransom RF, Vincent AM, et al. Induction of antioxidant enzymes in murine podocytes precedes injury by puromycin aminonucleoside. *Kidney Int* 2004; 66: 1881–1889.
248. Kojima K, Matsui K, Nagase M. Protection of alpha(3) integrin-mediated podocyte shape by superoxide dismutase in the puromycin aminonucleoside nephrosis rat. *Am J Kidney Dis* 2000; 35: 1175–1185.
249. de Vriese AS, Tilton RG, Elger M, et al. Antibodies against vascular endothelial growth factor improve early renal dysfunction in experimental diabetes. *J Am Soc Nephrol* 2001; 12: 993–1000.
250. Schrijvers BF, Flyvbjerg A, de Vriese AS. The role of vascular endothelial growth factor (VEGF) in renal pathophysiology. *Kidney Int* 2004; 65: 2003–2017.
251. Cooper ME, Vranes D, Youssef S, et al. Increased renal expression of vascular endothelial growth factor (VEGF) and its receptor VEGFR-2 in experimental diabetes. *Diabetes* 1999; 48: 2229–2239.
252. Wendt TM, Tanji N, Guo J, et al. RAGE drives the development of glomerulosclerosis and implicates podocyte activation in the pathogenesis of diabetic nephropathy. *Am J Pathol* 2003; 162: 1123–1137.
253. Iglesias-de la Cruz MC, Ziyadeh FN, Isono M, et al. Effects of high glucose and TGF-beta1 on the expression of collagen IV and vascular endothelial growth factor in mouse podocytes. *Kidney Int* 2002; 62: 901–913.
254. Chen S, Kasama Y, Lee JS, et al. Podocyte-derived vascular endothelial growth factor mediates the stimulation of alpha3(IV) collagen production by transforming growth factor-beta1 in mouse podocytes. *Diabetes* 2004; 53: 2939–2949.
255. Rizkalla B, Forbes JM, Cooper ME, Cao Z. Increased renal vascular endothelial growth factor and angiopoietins by angiotensin II infusion is mediated by both AT1 and AT2 receptors. *J Am Soc Nephrol* 2003; 14: 3061–3071.
256. Kakizawa H, Itoh Y, Imamura S, et al. Possible role of VEGF in the progression of kidney disease in streptozotocin (STZ)-induced diabetic rats: effects of an ACE inhibitor and an angiotensin II receptor antagonist. *Horm Metab Res* 2004; 36: 458–464.
257. Freeburg PB, Robert B, St John P, Abrahamson DR. Podocyte expression of hypoxia-inducible factor (HIF)-1 and HIF-2 during glomerular development. *J Am Soc Nephrol* 2003; 14: 927–938.
258. Susztak K, Ciccone E, McCue P, et al. Multiple metabolic hits converge on CD36 as novel mediator of tubular epithelial apoptosis in diabetic nephropathy. *PLoS Med* 2005; 2: e45.
259. Foster RR, Hole R, Anderson K, et al. Functional evidence that vascular endothelial growth factor may act as an autocrine factor on human podocytes. *Am J Physiol Renal Physiol* 2003; 284: F1263–F1273.
260. Ziyadeh FN. Mediators of diabetic renal disease: the case for TGF-beta as the major mediator. *J Am Soc Nephrol* 2004; 15(suppl 1): S55–S57.
261. Kopp JB, Factor VM, Mozes M, et al. Transgenic mice with increased plasma levels of TGF-beta 1 develop progressive renal disease. *Lab Invest* 1996; 74: 991–1003.
262. Ziyadeh FN, Hoffman BB, Han DC, et al. Long-term prevention of renal insufficiency, excess matrix gene expression, and glomerular mesangial matrix expansion by treatment with monoclonal antitransforming growth factor-beta antibody in db/db diabetic mice. *Proc Natl Acad Sci U S A* 2000; 97: 8015–8020.
263. Schiffer M, Mundel P, Shaw AS, Bottinger EP. A novel role for the adaptor molecule CD2-associated protein in transforming growth factor-beta-induced apoptosis. *J Biol Chem* 2004; 279: 37004–37012.
264. Suri C, Jones PF, Patan S, et al. Requisite role of angiopoietin-1, a ligand for the TIE2 receptor, during embryonic angiogenesis. *Cell* 1996; 87: 1171–1180.
265. Kanesaki Y, Suzuki D, Uehara G, et al. Vascular endothelial growth factor gene expression is correlated with glomerular neovascularization in human diabetic nephropathy. *Am J Kidney Dis* 2005; 45: 288–294.
266. Satchell SC, Anderson KL, Mathieson PW. Angiopoietin 1 and vascular endothelial growth factor modulate human glomerular endothelial cell barrier properties. *J Am Soc Nephrol* 2004; 15: 566–574.
267. Ferrara N. Vascular endothelial growth factor and the regulation of angiogenesis. *Recent Prog Horm Res* 2000; 55: 15–35.
268. Bowden DW. Genetics of diabetes complications. *Curr Diab Rep* 2002; 2: 191–200.
269. Baelde HJ, Eikmans M, Doran PP, et al. Gene expression profiling in glomeruli from human kidneys with diabetic nephropathy. *Am J Kidney Dis* 2004; 43: 636–650.
270. Mishra R, Emancipator SN, Miller C, et al. Adipose differentiation-related protein and regulators of lipid homeostasis identified by gene expression profiling in the murine db/db diabetic kidney. *Am J Physiol Renal Physiol* 2004; 286: F913–F921.
271. Wilson KH, Eckenrode SE, Li QZ, et al. Microarray analysis of gene expression in the kidneys of new- and post-onset diabetic NOD mice. *Diabetes* 2003; 52: 2151–2159.
272. Chung SS, Ho EC, Lam KS, Chung SK. Contribution of polyol pathway to diabetes-induced oxidative stress. *J Am Soc Nephrol* 2003; 14: S233–S236.
273. Chung SS, Chung SK. Aldose reductase in diabetic microvascular complications. *Curr Drug Targets* 2005; 6: 475–486.
274. Srinivasan V, Sandhya N, Sampathkumar R, et al. Glutamine fructose-6-phosphate amidotransferase

(GFAT) gene expression and activity in patients with type 2 diabetes: inter-relationships with hyperglycaemia and oxidative stress. *Clin Biochem* 2007; 40: 952–957.

275. Koschinsky T, He CJ, Mitsuhashi T, et al. Orally absorbed reactive glycation products (glycotoxins): an environmental risk factor in diabetic nephropathy. *Proc Natl Acad Sci U S A* 1997; 94: 6474–6479.
276. Uribarri J, Cai W, Sandu O, et al. Diet-derived advanced glycation end products are major contributors to the body's AGE pool and induce inflammation in healthy subjects. *Ann N Y Acad Sci* 2005; 1043: 461–466.
277. Negrean M, Stirban A, Stratmann B, et al. Effects of low-and high-advanced glycation endproduct meals on macro-and microvascular endothelial function and oxidative stress in patients with type 2 diabetes mellitus. *Am J Clin Nutr* 2007; 85: 1236–1243.
278. Stirban A, Negrean M, Stratmann B, et al. Adiponectin decreases postprandially following a heat-processed meal in individuals with type 2 diabetes: an effect prevented by benfotiamine and cooking method. *Diabetes Care* 2007; 30: 2514–2516.
279. Goldberg T, Cai W, Peppa M, et al. Advanced glycoxidation end products in commonly consumed foods. *J Am Diet Assoc* 2004; 104: 1287–1291.
280. Stirban A, Negrean M, Stratmann B, et al. Benfotiamine prevents macro-and microvascular endothelial dysfunction and oxidative stress following a meal rich in advanced glycation end products in individuals with type 2 diabetes. *Diabetes Care* 2006; 29: 2064–2071.
281. Bjelakovic G, Nikolova D, Gluud LL, et al. Mortality in randomized trials of antioxidant supplements for primary and secondary prevention: systematic review and meta-analysis. *JAMA* 2007; 297: 842–857.
282. Schmidt S, Ritz E. Genetic determinants of diabetic renal disease and their impact on therapeutic interventions. *Kidney Int Suppl* 1997; 63: S27–S31.

Past, Present, and Future of Insulin Pump Therapy: Better Shot At Diabetes Control

Jennifer Sherr, MD, and William V. Tamborlane, MD

Department of Pediatrics and Yale Center for Clinical Investigation, Yale University School of Medicine, New Haven, CT

ABSTRACT

With the advent of continuous subcutaneous insulin infusion therapy and the findings of the Diabetes Control and Complications Trial, the management of type 1 diabetes has changed drastically. Over the past 30 years since its development, the effectiveness of continuous subcutaneous insulin infusion has been assessed in comparison with other modes of intensive treatment. Additionally, improvements in pump delivery systems have been made. Here, the findings of the studies on pump therapy are reviewed. Selection criteria of patients for pump use and how to initiate pump therapy are presented. Finally, newer findings on continuous glucose sensors are discussed as the next era of pump therapy continues to focus on the goal of developing an artificial pancreas. ***Mt Sinai J Med 75:352–361, 2008.*** © *2008 Mount Sinai School of Medicine*

Key Words: continuous subcutaneous insulin infusion, insulin pumps, multiple daily injections, type 1 diabetes.

Continuous subcutaneous insulin infusion (CSII) pump therapy was introduced to treat patients with type 1 diabetes in the late 1970s.[1,2] Until then, the care of individuals with type 1 diabetes was crude, involving 1 or 2 daily injections of neutral protamine Hagedorn (NPH) and regular insulin of animal origin that was not purified, adjustments of insulin doses based on urinary glucose excretion, and dietary counseling focused on limiting simple sugars and maintaining fixed macronutrient intake at each meal. Because of the limitations of this regimen and fear of hypoglycemia, particularly in children, glucose levels often averaged over 300 mg/dL. It is also important to remember that at that time, a causal relationship between poor glycemic control and the development of complications was suspected but not yet proven. Thus, it is not surprising that the demonstration that CSII could provide a more physiologic method of insulin replacement in type 1 diabetes mellitus (T1DM) was enthusiastically received.[1–4] Self-monitoring of blood glucose (SMBG), hemoglobin A1c (HbA1c) assays, and purified insulin preparations were also introduced in the late 1970s, and the basal bolus approach used for CSII was adapted for use in multiple daily injection (MDI) regimens. These advances made intensive treatment of T1DM possible and set the stage for the Diabetes Control and Complications Trial (DCCT), which was launched in 1983.

Intensive therapy did not come without a price. An increased risk of severe hypoglycemia, due in part to reductions in counterregulatory hormone responses to hypoglycemia, was demonstrated.[5,6] This acquired inability to symptomatically identify low blood glucose levels due to repeated episodes of mild hypoglycemia was named *hypoglycemia-associated autonomic failure*[7] and made patients even more vulnerable to severe hypoglycemic events. Consequently, the need to solve the puzzle of the relationship between hyperglycemia and vascular complications became even more urgent and compelling.

Address Correspondence to:

Jennifer Sherr, MD
Department of Pediatrics
Yale University School of Medicine
New Haven, CT
Email: jennifer.sherr@yale.edu

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20055

© 2008 Mount Sinai School of Medicine

SETTING THE BAR: RESULTS OF THE DIABETES CONTROL AND COMPLICATIONS TRIAL

In 1993, the findings of the DCCT were published and showed that the benefits achieved by intensive therapy with respect to the development and progression of complications outweighed the associated 2- to 3-fold increase in the risk of severe hypoglycemia.[8] This beneficial effect was also shown in a subgroup analysis of the relatively small number of adolescents who participated in the study[9] and persisted for many years after completion of the randomized clinical trial.[10] With this focus on intensive therapy to prevent complications and improvements in pump technology, use of CSII increased dramatically, with the most rapid growth occurring in children and adolescents with T1DM who started to receive this therapy.

NONRANDOMIZED STUDIES OF PUMP THERAPY

Very few children and adolescents with T1DM used CSII before the late 1990s. Even though the percentage of intensively treated subjects who switched from MDI to CSII therapy increased steadily during the DCCT, the majority of these subjects were adults with T1DM. In an article[9] describing the analyses of DCCT results in the subset of adolescents, the investigators themselves questioned whether the goals of intensive therapy could be effectively translated into pediatric diabetes practice. To examine this question, the Adolescents Benefit from Control of Diabetes Study was undertaken by Grey and colleagues.[11] In this prospective study, 75 subjects, ranging in age from 12 to 20 years, were started on intensive therapy. Subjects chose the mode of intensive therapy delivery, with 25 choosing CSII and 50 choosing MDI. Those in the CSII group had a mean drop in HbA1c of 0.9%, whereas those in the MDI group had a mean drop of 0.5%. Despite this improvement in control, the CSII group demonstrated a 50% reduction in episodes of severe hypoglycemia, a finding that was the opposite of the DCCT data. However, weight gain increased with intensive therapy in both MDI-treated and CSII-treated subjects with better metabolic control. Interestingly, adolescents on CSII reported less difficulty in coping with diabetes.

Since that report, many clinical outcome studies have examined the effects of switching from MDI to CSII in children of various ages (Table I). A very consistent picture has emerged from these studies: mean HbA1c falls by 0.2% to 0.9%,[12–25] the frequency of clinically important hypoglycemia is reduced,[12–19,21–25] and body mass index z-scores do not increase.[12–16,18,20–25] Of note, the mean HbA1c values achieved across these studies (ie, ~7.5%) are considerably lower than those attained by adolescents in the intensive group in the DCCT (ie, 8.1%). In a 2003 meta-analysis of 11 studies with a parallel design, the weighted summary mean difference in HbA1c between CSII and MDI/conventional therapy was 0.95%, with a 95% confidence interval of 0.8% to 1.1%.[26]

From these nonrandomized studies, some additional observations deserve attention. In 2 studies, patients were switched from conventional therapy to either MDI with glargine (the current gold standard

Table 1. *Results of Switching from Injection Therapy to Continuous Subcutaneous Insulin Infusion Therapy in Nonrandomized Pediatric Studies.*

Authors	n	Age (Years)	Change in Hemoglobin A1c from Baseline % (*P*)	Hypoglycemia	Body Mass Index
Ahern *et al.*[12]	161	1–18	0.6–0.7 (<0.02)	Reduced	No change
Maniatis *et al.*[13]	56	7–23	0.2 (.045)	Reduced	No change
Sulli and Shashaj[14]	40	4–25	0.7 (<0.05)	Reduced	No change
Plotnick *et al.*[15]	95	4–18	0.4 (<0.001)	Reduced	No change
Willi *et al.*[16]	51	1–16	0.45 (<0.01)	Reduced	No change
Alemzadeh *et al.*[17]	40	10–18	0.6 (<0.002)	Reduced	Increased
Weinzimer *et al.*[18]	65	1–6	0.6 (0.003)	Reduced	Slight decrease
Mack-Fogg[19]	70	2–12	0.5 (<0.0001)	Reduced	Slight increase
Schiaffini *et al.*[20]	20	6–18	0.9 (<0.05)	No change	No change
Jeha *et al.*[21]	10	1–6	0.9 (0.01)	Reduced	No change
Nimri *et al.*[22]	279	1–40	0.51 (<0.01)	Reduced	No change
Behre *et al.*[23]	33	2–7	0.7 (<0.001)	Reduced	No change
Sulli and Shashaj[24]	42	4–17	0.7 (0.00)	Reduced	Reduced
Scrimgeour *et al.*[25]	291	Mean age: 13.3	0.4 (<0.0001)	Reduced	No change

of MDI treatment) or CSII.[27] Subjects who switched to CSII showed improved glycemic control versus those who switched to glargine.[17,20] Concern also existed about the use of CSII in infants, toddlers, and preschool children, but a number of studies have shown that CSII is particularly efficacious in this age group.[12,18,21,23] Other studies have highlighted that CSII may be beneficial in subjects with the highest baseline HbA1c, with a mean decrease in HbA1c of 1.7% in the subpopulation of patients who had a baseline HbA1c greater than 10% in Nimri *et al.*'s study.[22] It is also important to realize that the beneficial effects of CSII were sustained during long-term use in 3 of the 4 nonrandomized studies that followed patients for more than 3.5 years.[18,24,25,28]

The use of insulin pump therapy from the time of diagnosis has also been investigated. In the early days of pump therapy, a 2-year randomized clinical trial evaluated the efficacy of CSII in comparison with conventional therapy in 30 children with new-onset T1DM and demonstrated that the CSII arm had significantly lower HbA1c levels.[29] It was interesting that CSII was well accepted in newly diagnosed patients at a time when such therapy was still in its infancy. In a more recent nonrandomized study, 28 patients were offered pump therapy within the first month of diagnosis, and all accepted participation in the study.[30] As expected, improved glycemic control was seen with CSII. Equally important, patients reported no adverse impacts of diabetes on the flexibility of their lifestyle in comparison with their lifestyle before diagnosis, and this showed the beneficial effects from a psychosocial standpoint with respect to pump therapy.[30]

A recent database review of clinical factors involved in HbA1c levels, including gender, ethnicity, socioeconomic status, duration of diabetes, and insulin regimen, was reported by our clinic.[31] One of the strongest predictors of low HbA1c was found to be use of CSII, with 286 patients on this therapy having a mean HbA1c of 7.2% versus 8.1% in the 167 patients on MDI therapy.[31] In the same study, it was demonstrated that lower socioeconomic status was associated with poor metabolic control.[31] Clearly, nonrandomized studies have paved the path for pump therapy, which has been supported further by randomized trials.

RANDOMIZED STUDIES OF PUMP THERAPY

A number of randomized studies have assessed the efficacy of CSII versus MDIs with NPH as the basal insulin (Table 2). DeVries and colleagues[32] reported on 79 patients recruited from 11 Dutch centers in a randomized crossover design. Given a high dropout rate, the trial was transitioned to a parallel clinical trial, which showed improved HbA1c in those on CSII therapy. A randomized, crossover trial was completed by Weintrob and colleagues[33] in which 23 children were studied on both MDI and CSII for a duration of 3.5 months for each therapy. This study demonstrated no difference between treatment groups with respect to HbA1c or hypoglycemia; however, the patients reported higher treatment satisfaction on CSII. The large Five Nations Trial demonstrated lower HbA1c, less blood glucose variability, and higher quality of life scores in those receiving CSII therapy.[34] Other studies have demonstrated no difference in HbA1c in MDI versus CSII,[35–38] but the majority of these studies were in very young children, whose target glucose levels were set considerably higher than those of older children and adolescents. Moreover, it is very clear that a major advantage of CSII in the youngest pediatric patients from the perspective of parents is that it makes living with diabetes much more bearable.[39]

A meta-analysis assessing CSII versus intensive insulin injection therapy via randomized, controlled trials was published in 2002. This meta-analysis examined 12 randomized, controlled trials and showed a reduction of 0.44 in HbA1c (confidence interval: 0.20–0.69) in patients using CSII.[40] There was also an approximately 14% drop in total daily insulin dose in patients using CSII.[40] Many of the studies in this meta-analysis used NPH rather than long-acting insulin analogues in their MDI groups, even though the peaking of NPH and its variable absorption make it less ideal for basal insulin replacement than glargine or detemir.[27]

Table 2. *Results of Randomized Clinical Trials Comparing CSII and MDI in Patients with Type 1 Diabetes Mellitus.*

Authors	n	Age (Years)	A1c: CSII Versus MDI (%)
De Vries *et al.*[32]	79	18–70	8.4 versus 9.2*
Weintrob *et al.*[33]	23	9–14	8.0 versus 7.9
Hoogma *et al.*[34]	279	18–65	7.4 versus 7.7*
DiMeglio *et al.*[35]	42	<5	8.5 versus 8.7
Wilson *et al.*[36]	19	1–7	7.8 versus 8.1
Fox *et al.*[37]	26	1–6	7.2 versus 7.5
Opipari-Arrigan *et al.*[38]	16	3–6	8.4 versus 8.2
Doyle *et al.*[41]	32	8–21	7.2 versus 8.1*
Schiaffini *et al.*[42]	36	9–18	7.6 versus 8.2*

Abbreviations: CSII, continuous subcutaneous insulin infusion; MDI, multiple daily injection.
* $P < 0.05$.

A randomized, parallel group clinical trail was completed by our group in 32 conventionally treated subjects who were randomized to either CSII or glargine and bolus doses of aspart.[41] The CSII group was found to have lower HbA1c levels in comparison with the glargine group and also in comparison with their baseline HbA1c levels. Fasting plasma glucose concentrations did not differ between the 2 groups, indicating similar adequacy of basal insulin replacement. However, SMBG performed at lunch, dinner, and bedtime demonstrated significantly lower blood glucose levels in the CSII group.[41] These findings suggest that lack of compliance with bolus dosing likely contributed to the higher blood glucose and HbA1c levels in the MDI group. Schiaffini and colleagues[42] demonstrated similar findings in their randomized trial of 36 children between 9 and 18 years of age who were followed for 2 years after being randomized to either MDI with glargine or CSII.[42] Although both treatment arms showed improvement in glycemic control based on HbA1c levels at 1 year, only the CSII group continued to demonstrate improved HbA1c after 2 years of therapy.[42]

The very flat time-action profiles of new long-acting insulin analogs place a premium on strict compliance with and the need for insulin injections for every meal and snack, which may be difficult for older children and adolescents. This is a reason that several pediatric diabetes centers use a compromise approach that employs NPH mixed with aspart or lispro in the morning and long-acting analogues with aspart or lispro at dinner.[43] Some even mix the rapid and long-acting analogs together to avoid an extra injection without adverse effects on metabolic control.[44]

BOLUS AND BEYOND: THE NEW AND IMPROVED INSULIN PUMPS

While many of the studies demonstrating the effectiveness of pump therapy were being completed, insulin analogues and technological advances changed the way in which CSII therapy evolved. The desire for smaller, more portable pumps[45,46] was realized as pumps shrank to the size of a pager. In the DCCT, regular insulin was used; however, the advantages of rapidly acting insulin analogues (lispro and aspart) are now well recognized.[47,48] With regular insulin, the delayed peak leads to postprandial hyperglycemia, whereas rapidly acting analogues allow for a more physiologic peak in insulin action. As the duration of action of regular insulin is longer, the risk of hypoglycemia in the late postprandial period is also increased. It should be noted, however, that the goal of mimicking the natural peak and duration of endogenous insulin remains elusive even with rapidly acting analogues.[49]

The technological capabilities of pumps have also advanced. Programmable pumps have allowed for titration of basal rates and varying bolus doses with a wide array of delivery options. Programmable pumps have been shown to lead to improved preprandial glycemic control and fewer episodes of overnight hypoglycemia versus nonprogrammable pumps.[50] One of the features of pumps that is particularly important in treating adolescents with T1DM is the bolus history function, which allows clinicians and parents to assess whether patients are missing bolus doses of insulin.[51,52] In 48 patients on pump treatment, a cross-sectional study demonstrated that those who missed less than 1 bolus per week versus those who missed 1 or more mealtime boluses had better HbA1c: 8.0% versus 8.8%, respectively ($P = 0.0001$).[52] Mealtime alarms can be set with CSII, which may provide at least a temporary improvement in metabolic control.[53]

The fear of hypoglycemia remains a major obstacle that prevents patients and parents from achieving HbA1c goals. Although patients with T1DM are encouraged to exercise regularly, the risk of hypoglycemia during and on the night after exercise is substantially increased. Recent studies completed by the Diabetes Research in Children Network showed that the risk of hypoglycemia increased both during exercise and on the night following a 75-minute period of moderate-intensity aerobic exercise in patients on fixed insulin dosing.[54,55] A randomized crossover study was subsequently performed to assess the efficacy of suspending the basal rate on CSII during exercise. Suspending basal rates decreased the risk of hypoglycemia, defined as blood glucose <70 mg/dL, from 43% to 16%.[56] The ability to individually titrate the correct basal rate not only during exercise but also on the night following exercise provides a clear benefit for those on CSII versus those on fixed MDI regimens.[57]

With all of the advances in pump therapy, the question of when youth with type 1 diabetes can begin to have autonomy with their care has been debated. One study showed that parents' reports of CSII skill mastery and the age at which clinicians expect skill mastery are similar.[58] Yet, the keys to a successful transition remain adult supervision during the transition period and tailoring the timing of autonomy on the basis of an individual's maturity and understanding of pump therapy.

GOING BEYOND THE NUMBERS: PSYCHOSOCIAL BENEFITS OF CONTINUOUS SUBCUTANEOUS INSULIN INFUSION

Nonrandomized and randomized studies have shown benefits of CSII that go beyond improvements in glycemic control. Quality of life scores measured via different scales have shown improvement in patients treated with CSII, and CSII-treated patients have reported higher treatment satisfaction than those on MDI therapy.[32–34] A large nonrandomized case-control study recently showed greater lifestyle flexibility, decreased fear of hypoglycemia, and improved treatment satisfaction when CSII subjects were compared to those on MDI therapy, whether it was a glargine-based or NPH-based regimen.[59] In a study of 16 children 3 to 5 years old, parents reported decreased diabetes-related worry in patients on CSII.[41] In 3 randomized studies, more than 95% of subjects decided to continue on CSII therapy after study completion.[35–37]

In a recent study, parents of young children with T1DM were found to have a moderate level of fear with respect to hypoglycemia, and there was a positive correlation between fear and mean daily blood glucose levels.[60] Fear of hypoglycemia may be a particular problem in young children in day care. However, in a study of children under the age of 7 being treated with CSII, HbA1c levels and risk of severe hypoglycemia were lower when children were cared for by paid providers than when the mother was the primary daytime caregiver.[18] In a qualitative study, parents of young children with T1DM reported that "everyone in the family experienced more freedom, flexibility, and spontaneity in their daily lives" with CSII therapy.[39] A therapy that provides better disease management while allowing for greater normalcy in life is ideal for any chronic disease in pediatrics, and CSII therapy clearly represents the best current approximation of this ideal to many families.

PRICE OF IMPROVED CARE

In the DCCT, the cost of intensive care was assessed, and it was determined that MDI therapy cost approximately $4000/year, whereas CSII therapy cost $5800/year.[61] This was in comparison with conventional therapy, which was estimated to cost $1700/year.[61] Most of the cost in the intensive treatment group was associated with a greater frequency of outpatient visits and resources used. The additional cost of CSII therapy was associated with the cost of the pump and supplies required for the pump. However, in light of the decreased risk of complications associated with improved glycemic control, the benefits of intensive therapy would be expected to have an effect on later healthcare costs associated with treatment for complications. In a retrospective analysis of annual retinopathy screening in our practice, none of the patients who met American Diabetes Association screening guidelines for retinopathy screening were found to have retinopathy.[62] Therefore, with more stringent control provided by intensive therapy, the guidelines for routine screening of complications may need to be revised, and costs for both the treatment of complications and the screening of complications may be reduced.

Reimbursement for the preventative services provided to youth with T1DM covers only a fraction of the costs associated with multidisciplinary team management of this disease.[63] At our center, advanced nurse practitioners are key members of the team who interact most frequently with the patients and parents.[64] However, lack of reimbursement for nonphysician members of the treatment team has led to financial problems.[63] Therefore, although pump therapy may remain more expensive and intensive therapy requires more involvement of the multidisciplinary team, the risk of complication development would pose a much greater burden on healthcare costs.

PRACTICAL CONSIDERATIONS OF PUMP THERAPY

On the basis of the findings of randomized and nonrandomized studies, it is safe to conclude that CSII is an effective method for the treatment of patients with type 1 diabetes, regardless of age. This sentiment has been echoed in a position statement by the Lawson-Wilkins Drug and Therapeutics Committee and in the recommendations of a recent consensus conference that examined the evidence supporting the use of pumps in pediatrics.[65,66] CSII was deemed to be the most physiologic method of insulin delivery currently available. It was also noted that CSII offers the possibility of more flexibility and more precise insulin delivery than MDI.[66] Guidance for when CSII therapy should be considered was also provided in the consensus statement (Table 3). Moreover, the consensus conference indications for using pumps in pediatrics cover virtually any child or adolescent with T1DM. However, it is just as important that treatment teams select appropriate candidates for this therapy.

Table 3. *Indications for Use of CSII in Pediatrics.*

Conditions Under Which CSII Should Be Considered
1. Recurrent severe hypoglycemia
2. Wide fluctuations in blood glucose levels regardless of A1c
3. Suboptimal diabetes control (ie, A1c exceeds target range for age)
4. Microvascular complications and/or risk factors for macrovascular complications
5. Good metabolic control but insulin regimen that compromises lifestyle
Circumstances in Which CSII May Be Beneficial
1. Young children and especially infants and neonates
2. Adolescents with eating disorders
3. Children and adolescents with a pronounced dawn phenomenon
4. Children with needle phobia
5. Pregnant adolescents, ideally before conception
6. Ketosis-prone individuals
7. Competitive athletes

NOTE: This table was adapted from ref. 66.
Abbreviation: CSII, continuous subcutaneous insulin infusion.

In order to be considered for CSII in our clinic, patients must be performing at least 4 blood glucose measurements per day, consistently attend follow-up visits, be committed to the goals of intensive insulin therapy, have a solid understanding of basic diabetes management (including carbohydrate counting), and be in at least fair control (ie, HbA1c $\leqslant$ 8.5%), which is used as an index of compliance with current injection therapy. Once a patient is deemed appropriate for CSII therapy, the selection of a pump is necessary. Table 4 reviews various pumps currently available for use and some of the features of these pumps. All of these pumps are considered "smart" and calculate the dose of insulin required for carbohydrate coverage or the correction dose of insulin. The selection of a pump is a based on features desired by the patient/family along with guidance from the multidisciplinary team.

Families are encouraged to review educational materials, DVDs, and computer programs that are shipped with the pumps. Following this, an outpatient visit lasting approximately 60 to 90 minutes is completed to start the pump. Dosages for pump initiation are based on the current total daily dose of insulin that the patient is receiving. Approximately 50% of the total daily dose is given as basal insulin. This basal insulin is divided across the 24-hour period. For a patient receiving 40 units/day prior to pump therapy, 20 units would be the total daily basal insulin dosage, and this would be given as approximately 0.8 units/hour. Various methods are employed to determine correction factors and carbohydrate ratios. If someone is already using these as part of their injection therapy, then these same factors can be carried forward. Alternatively, the carbohydrate ratio can be calculated by the division of 500 by the total daily dose. Similarly, a correction factor can be derived by the division of 1800 by the total daily dose. An even simpler approach that we often use is to determine the carbohydrate-to-insulin ratios and correction factors on the basis of the patient's age(Table 5).[67]

Patients are advised to check their blood glucose before meals, at bedtime, at 12 a.m., and at 3 a.m. and call the center as needed for dose adjustments. After pump initiation, patients return to routine follow-up, being seen in the clinic every 3 months, with dose adjustments and an emergency line available as needed.

CLOSING THE LOOP: ADVANCES IN CONTINUOUS GLUCOSE MONITORING

Continuous glucose monitoring technology has the potential to revolutionize the treatment of T1DM. In comparison with SMBG, which gives the patient only brief snapshots of where their blood glucose lies at the time of the test, continuous glucose sensors provide the opportunity to look at streaming videos of data, on the basis of which changes in dosage can be made. Nevertheless, the ability of currently available devices to provide tangible benefits to patients and their families remains to be established.

In a study of metabolic control in 56 children and adolescents with T1DM that used the Minimed continuous glucose monitoring system, almost 90% of patients had a peak postprandial glucose level greater than 180 mg/dL after every meal despite HbA1c levels averaging 7.7%, and 70% of patients had a period of asymptomatic hypoglycemia (defined as glucose <60 mg/dL).[68] Although the original continuous glucose monitoring system provided insights into management problems, the clinical utility of this retrospective, Holter-type system in youth with T1DM is limited.

The GlucoWatch Biographer was the first real-time continuous glucose monitor (RT-CGM) to be introduced, but initial studies involving this device were very disappointing. In a large randomized clinical trial comparing the GlucoWatch Biographer and standard SMBG, no change in HbA1c was noted in either group, and patients' use of the device

Table 4. *Pump Options and Features.*

Pump	Insulin Reservoir Capacity (U)	Minimal Basal Rate Increment (U/hour)	Minimal Bolus Dose Increment (U)	Other Features
Animas IR-2020	200	0.025	0.05	• Smallest pump • Largest display screen • 500-food individualized database
Deltec Cozmo	300	0.05	0.05	• Integrated freestyle meter • Enhanced meal maker • Basal rates by day of week • Replacement of basal rate after disconnection of pump
Disetronic Spirit	315	0.1	0.1	• Reversible display • Menu display customization option
Medtronic Paradigm 522/722	180 or 300	0.05	0.1	• Only available pump with real-time CGMS on the market • Optional remote control for bolus dosing • CareLink personal therapy management tool
Insulet Omnipod	200	0.05	0.05	• No tubing • 1000 common foods in PDA • Freestyle meter in the PDA component

Abbreviation: CGMS, continuous glucose monitoring system; PDA, personal digital assistant.

Table 5. *Correction Factors and Carbohydrate Ratios Based on Age for Initial Pump Settings.*

Age (Years)	Correction Factor (mg/dL)	Carbohydrate Ratio (g)*
<3	225	45
4–5	200	40
5–7	150	30
8–11	125	20
Prepubertal and/or <13	75	15
Pubertal and/or>13	50	10

*Often a lower carbohydrate ratio is needed for breakfast (−2 g for the breakfast carbohydrate ratio). For example, for children less than 3 years old, the breakfast carbohydrate ratio should be started at 43 g.

declined rapidly because of skin irritation, excessive alarms, and inaccurate readings.[69]

The current generation of RT-CGM devices being manufactured by DexCom, Medtronic, and Abbott are more accurate and user friendly than the first generation of continuous glucose monitor systems,[70–72] and early studies have suggested that they may be beneficial in the management of youth with T1DM.[73,74] A large-scale randomized clinical trial of all 3 of these systems is currently in progress and should provide a much more comprehensive assessment of their benefits and limitations.

With all the advances in pump technology and improvements in glycemic control, the next frontier for pump therapy will involve closing the loop and making the dream of an artificial pancreas a reality. Indeed, automated closed-loop insulin delivery systems that combine external insulin pumps with current RT-CGM devices are already being studied. In a study of 10 patients using a completely automated insulin delivery system, the amount of time that the blood glucose was between 70 and 180 mg/dL rose from 63% to 75%.[75] However, an elevation in postprandial blood glucose led to the hypothesis that a priming bolus of insulin prior to meals could aid in preventing these meal-related glycemic excursions. This theory has subsequently been tested by a comparison of 8 patients on fully automated pumps versus 9 patients on a hybrid closed-loop system (premeal priming bolus with a closed-loop system).[76] Peak postprandial glucose levels were significantly lower in patients on the hybrid system versus those on the fully automated system.[76] With continued study and manipulation, it appears that the artificial pancreas could become a viable treatment option for patients with T1DM in the foreseeable future.

CONCLUSION

CSII therapy has revolutionized diabetes care. With the findings of the DCCT, the importance of stringent glycemic control was identified. Since that time, randomized and nonrandomized studies have shown the efficacy of CSII across all age groups.

Continuous glucose sensors are now changing the way in which CSII therapy can improve control by reducing glycemic excursions and manipulating insulin delivery to avoid otherwise asymptomatic hypoglycemia detected by these sensors. The endeavor of developing an artificial pancreas is transitioning from a dream to reality as newer studies have demonstrated its use and effectiveness in small cohorts.

ACKNOWLEDGMENT

This research was supported by grants from the Stephen J. Morse Pediatric Diabetes Research Fund, the Juvenile Diabetes Research Foundation, and the National Institutes of Health (T32 DK063703).

DISCLOSURES

Potential conflict of interest: William V. Tamborlane is a member of the advisory boards for Medtronic, Abbott Diabetes Care, and Novo Nordisk and a member of the speaker's bureaus for Medtronic and Novo Nordisk.

REFERENCES

1. Tamborlane WV, Sherwin RS, Genel M, Felig P. Reduction to normal of plasma glucose in juvenile diabetics by subcutaneous administration of insulin with a portable infusion pump. *N Engl J Med* 1979; 300: 573–578.
2. Pickup JC, Keen H, Parsons JA, Alberti KG. Continuous subcutaneous insulin infusion: an approach to achieving normoglycemia. *Br Med J* 1978; 1: 204–207.
3. Home PD, Pickup JC, Keen H, et al. Continuous subcutaneous insulin infusion: comparison of plasma insulin profiles after infusion or bolus injection of the mealtime dose. *Metabolism* 1981; 30: 439–442.
4. Lauritezen T, Pramming S, Deckert T, Binder C. Pharmacokinetics of continuous subcutaneous insulin infusion. *Diabetologia* 1983; 24: 326–329.
5. Amiel SA, Tamborlane WV, Sherwin RS. Defective glucose counterregulation after strict glycemic control of insulin dependent diabetes mellitus. *N Engl J Med* 1987; 316: 1376–1383.
6. Amiel SA, Sherwin RS, Simonson DC, Tamborlane WV. Effect of intensive insulin therapy on glycemic thresholds for counterregulatory hormone release. *Diabetes* 1988; 37: 901–907.
7. Cryer PE. Iatrogenic hypoglycemia as a cause of hypoglycemia-associated autonomic failure in IDDM: a vicious cycle. *Diabetes* 1992; 41: 255–260.
8. DCCT Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin dependent diabetes mellitus. The Diabetes Control and Complications Trial. *N Engl J Med* 1993; 329: 977–986.
9. DCCT Research Group. The effect of intensive diabetes treatment on the development and progression of long-term complications in adolescents with insulin-dependent diabetes mellitus: the Diabetes Control and Complications Trial. *J Pediatr* 1994; 125: 177–188.
10. DCCT/EDIC Research Group. Beneficial effects of intensive therapy of diabetes during adolescence outcomes after the conclusion of the Diabetes Control and Complications Trial. *J Pediatr* 2001; 139: 804–812.
11. Boland, EA, Grey M, Oesterle A, et al. Continuous subcutaneous insulin infusion: a new way to achieve strict metabolic control, decrease severe hypoglycemia and enhance coping in adolescents with type 1 diabetes. *Diabetes Care* 1999; 22: 1779–1784.
12. Ahern JA, Boland EA, Doane R, et al. Insulin pump therapy in pediatrics: a therapeutic alternative to safely lower HbA1c levels across all age groups. *Pediatr Diabetes* 2002; 3: 10–15.
13. Maniatis AK, Klingensmith GJ, Slover RH, et al. Continuous subcutaneous insulin infusion therapy for children and adolescents: an option for routine diabetes care. *Pediatrics* 2001; 107: 351–356.
14. Sulli N, Shashaj B. Continuous subcutaneous insulin infusion in children and adolescents with diabetes mellitus: decreased HbA1c with low risk of hypoglycemia. *J Pediatr Endocrinol* 2003; 16: 393–399.
15. Plotnick LP, Clark LM, Brancati FL, Erlinger T. Safety and effectiveness of insulin pump therapy in children and adolescents with type 1 diabetes. *Diabetes Care* 2003; 26: 1142–1146.
16. Willi SM, Planton J, Egede L, Schwarz S. Benefits of continuous subcutaneous insulin infusion in children with type 1 diabetes mellitus. *J Pediatr* 2003; 143: 796–801.
17. Alemzadeh, R, Ellis JN, Holzum MK, et al. Beneficial effects of continuous subcutaneous insulin infusion and flexible multiple daily insulin regimen using insulin glargine in type 1 diabetes. *J Pediatr* 2004; 114: e91–e95.
18. Weinzimer SA, Ahern JA, Doyle EA, et al. Persistence of benefits of continuous subcutaneous insulin infusion in very young children with type 1 diabetes: a follow-up report. *J Pediatr* 2004; 114: 1601–1604.
19. Mack-Fogg JE, Orlowski CC, Jospe N. Continuous subcutaneous insulin infusion in toddlers and children with type 1 diabetes mellitus is safe and effective. *Pediatr Diabetes* 2005; 6: 17–21.
20. Schiaffini R, Ciampalini P, Spera S, et al. An observational study comparing continuous subcutaneous insulin infusion (CSII) and insulin glargine in children with type 1 diabetes. *Diabetes Metab Res Rev* 2005; 21: 347–352.
21. Jeha GS, Karaviti LP, Anderson B, et al. Insulin pump therapy in preschool children with type 1 diabetes mellitus improves glycemic control and decreases glucose excursions and the risk of hypoglycemia. *Diabetes Technol Ther* 2005; 7: 876–884.
22. Nimri R, Weintrob N, Benzaquen H, et al. Insulin pump therapy in youth with type 1 diabetes: a retrospective paired study. *J Pediatr* 2006; 117: 2126–2131.
23. Berhe T, Postellon D, Wilson B, Stone R. Feasibility and safety of insulin pump therapy in children aged 2 to 7 years with type 1 diabetes: a retrospective study. *J Pediatr* 2006; 117: 2132–2137.

24. Sulli N, Shashaj B. Long-term benefits of continuous subcutaneous insulin infusion in children with type 1 diabetes: a 4-year follow-up. *Diabet Med* 2006; 23: 900–906.
25. Scrimgeour L, Cobry E, McFann K, et al. Improved glycemic control after long-term insulin pump use in pediatric patients with type 1 diabetes. *Diabetes Technol Ther* 2007; 9: 421–428.
26. Weissberg-Benchell J, Antisdel-Lomaglio J, Seshadri R. Insulin pump therapy: a meta-analysis. *Diabetes Care* 2003; 26: 1079–1087.
27. Lepore M, Pampanelli S, Fanelli C, et al. Pharmacokinetics and pharmacodynamics of subcutaneous injection of long-acting human insulin analog glargine, NPH insulin, and ultralente human insulin, and continuous subcutaneous infusion of insulin lispro. *Diabetes* 2000; 49: 2142–2148.
28. Hanas R, Adolfsson P. Insulin pumps in pediatric routine care improve long-term metabolic control without increasing the risk of hypoglycemia. *Pediatr Diabetes* 2006; 7: 25–31.
29. De Beufort CE, Houtzagers CM, Bruinino GJ, et al. Continuous subcutaneous insulin infusion (CSII) versus conventional injection therapy in newly diagnosed diabetic children: two year follow up of a randomized prospective trial. *Diabet Med* 1989; 6: 766–771.
30. Ramachandani, N, Ten S, Anhalt H, et al. Insulin pump therapy from the time of diagnosis of type 1 diabetes. *Diabetes Technol Ther* 2006; 8: 663–670.
31. Springer D, Dziura J, Tamborlane WV, et al. Optimal control of type 1 diabetes mellitus in youth receiving intensive treatment. *J Pediatr* 2006; 149: 227–232.
32. DeVries J, Snoek F, Kostense P, Masurel N. A randomized trial of continuous subcutaneous insulin infusion and intensive injection therapy in type 1 diabetes for patients with long-standing poor glycemic control. *Diabetes Care* 2002; 25: 2074–2080.
33. Weintrob, N, Benzaquen H, Galatzer A, et al. Comparison of continuous subcutaneous insulin infusion and multiple daily injection regimens in children with type 1 diabetes: a randomized open crossover trial. *Pediatrics* 2003; 112: 559–564.
34. Hoogma R, Hammon P, Gomist R, et al., for the 5-Nations Study Group. Comparison of the effects of continuous subcutaneous insulin infusion (CSII) and NPH-based multiple daily insulin injections (MDI) on glycaemic control and quality of life: results of the 5 Nations Trial. *Diabet Med* 2006; 23: 141–147.
35. DiMeglio L, Pottorff T, Boyd S, et al. A randomized, controlled study of insulin pump therapy in diabetic preschoolers. *J Pediatr* 2004; 145: 380–384.
36. Wilson D, Buckingham B, Kunzelman E, et al. A two-center randomized controlled feasibility trial of insulin pump therapy in young children with diabetes. *Diabetes Care* 2005; 28: 15–19.
37. Fox L, Buckloh L, Smith S, et al. A randomized controlled trial of insulin pump therapy in young children with type 1 diabetes. *Diabetes Care* 2005; 28: 1277–1281.
38. Opipari-Arrigan L, Fredericks EM, Burkhart N, et al. Continuous subcutaneous insulin infusion benefits quality of life in preschool-age children with type 1 diabetes mellitus. *Pediatr Diabetes* 2007; 8: 377–383.
39. Sullivan-Bolyai S, Knafl K, Tamborlane W, Grey M. Parents' reflections on managing their children's diabetes with insulin pumps. *J Nurs Sch* 2004; 36: 316–323.
40. Pickup J, Matock M, Kerry S. Glycemic control with continuous subcutaneous insulin infusion compared with intensive insulin injections in patients with type 1 diabetes: meta-analysis of randomized controlled trials. *BMJ* 2002; 324: 1–6.
41. Doyle E, Weinzimer S, Steffen A, et al. A randomized, prospective trail comparing the efficacy of continuous subcutaneous insulin infusion with multiple daily injections using insulin glargine. *Diabetes Care* 2004; 27: 1554–1558.
42. Schiaffini R, Patera P, Bizzarri C, et al. Basal insulin supplementation in type 1 diabetic children: a long term comparative observational study between continuous subcutaneous insulin infusion and glargine insulin. *J Endocrinol Invest* 2007; 30: 572–577.
43. Chase HP, Dixon B, Pearson J, et al. Reduced hypoglycemic episodes and improved glycemic control in children with type 1 diabetes using insulin glargine and neutral protamine Hagedorn insulin. *J Pediatr* 2003; 143: 737–740.
44. Kaplan W, Rodriguez L, Smith O, et al. Effects of mixing glargine and short-acting insulin analogs on glucose control. *Diabetes Care* 2004; 27: 2739–2740.
45. Tamborlane W, Sherwin R, Genel M, Felig P. Outpatient treatment of juvenile-onset diabetes with a preprogrammed portable subcutaneous insulin infusion system. *Am J Med* 1980; 68: 190–196.
46. Pickup J, Keen H, Viberti G, Bious R. Patient reactions to long-term outpatient treatment with continuous subcutaneous insulin infusion. *Br Med J* 1981; 282: 766–768.
47. Renner R, Pfutzner A, Trautmann M, et al., for the German Humalog CSII Study Group. Use of insulin lispro in continuous subcutaneous insulin infusion treatment. *Diabetes Care* 1999; 22: 784–788.
48. Weinzimer S, Ternand C, Howard C, et al., for the Insulin Aspart Pediatric Pump Study Group. A randomized trial comparing continuous subcutaneous insulin infusion of insulin aspart versus insulin lispro in children and adolescents with type 1 diabetes. *Diabetes Care* 2008; 31: 210–215.
49. Swan, K, Weinzimer S, Dziura J, et al. Effect of puberty on the pharmacodynamic and pharmacokinetic properties of insulin pump therapy in youth with type 1 diabetes. *Diabetes Care* 2008; 31: 44–46.
50. Catargi B, Breilh D, Gin H, et al. A randomized study comparing blood glucose control and risk of severe hypoglycemia achieved by non-programmable versus programmable external insulin pumps. *Diabetes Metab* 2001; 27: 323–327.
51. Pankoska E, Skorka A, Szypowska A, Lipka M. Memory of insulin pumps and their record as a source of information about insulin therapy in children and adolescents with type 1 diabetes. *Diabetes Technol Ther* 2005; 7: 308–314.
52. Burdick J, Chase H, Slover R, et al. Missed insulin meal boluses and elevated hemoglobin A1c levels in children receiving insulin pump therapy. *Pediatrics* 2004; 113: e221–e224.
53. Chase H, Horner B, McFann K, et al. The use of insulin pumps with meal bolus alarms in children with type 1 diabetes to improve glycemic control. *Diabetes Care* 2006; 29: 1012–1015.

54. Diabetes Research in Children Network Study Group. Impact of exercise on overnight glycemic control in children with type 1 diabetes. *J Pediatr* 2005; 147: 528–534.
55. Diabetes Research in Children Network Study Group. The effects of aerobic exercise on glucose and counterregulatory hormone concentrations in children with type 1 diabetes. *Diabetes Care* 2006; 29: 20–25.
56. Diabetes Research in Children Network Study Group. Prevention of hypoglycemia during exercise in children with type 1 diabetes by suspending basal insulin. *Diabetes Care* 2006; 29: 2200–2204.
57. Tamborlane W. Triple jeopardy: nocturnal hypoglycemia after exercise in the young with diabetes. *JCEM* 2007; 92: 815–816.
58. Weissberg-Benchell J, Goodman S, Antisdel Lomaglio J, Zebracki K. The use of continuous subcutaneous insulin infusion (CSII): parental and professional perceptions of self-care mastery and autonomy in children and adolescents. *J Pediatr Psychol* 2007; 32: 1196–1202.
59. Nicoucci A, Maione A, Franciosis M, et al. Quality of life and treatment satisfaction in adults with type 1 diabetes: a comparison between continuous subcutaneous insulin infusion and multiple daily injections. *Diabet Med* 2008; 25: 213–220.
60. Patton S, Dolan L, Henry R, Powers S. Parental fear of hypoglycemia: young children treated with continuous subcutaneous insulin infusion. *Pediatr Diabetes* 2007; 8: 362–368.
61. DCCT Research Group. Resource utilization and costs of care in the diabetes control and complications trial. *Diabetes Care* 1995; 18: 1468–1478.
62. Huo B, Steffen A, Swan K, et al. Clinical outcomes and cost-effectiveness of retinopathy screening in youth with type 1 diabetes. *Diabetes Care* 2007; 30: 362–363.
63. Melzer S, Richards G, Covington M. Reimbursement and costs of pediatric ambulatory diabetes care by using the resource-based relative value scale: is multidisciplinary care financially viable? *Pediatr Diabetes* 2004; 5: 133–142.
64. Ahern J, Ramcahndani N, Cooper J, et al. Using a primary nurse manager to implement DCCT recommendations in a large pediatric program. *Diabetes Educ* 2000; 26: 990–994.
65. Eugster E, Francis G, and the Lawson-Wilkins Drug and Therapeutics Committee. Position statement: continuous subcutaneous insulin infusion in very young children with type 1 diabetes. *Pediatrics* 2006; 118: e1244–e1249.
66. Phillip M, Battelino T, Rodriguez H, et al. Use of insulin pump therapy in the pediatric age-group: consensus statement from the European Society for Paediatric Endocrinology, the Lawson Wilkins Pediatric Endocrine Society, and the International Society for Pediatric and Adolescent Diabetes, endorsed by the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care* 2007; 30: 1653–1662.
67. Tamborlane W, Fredrickson L, Ahern J. Insulin pump therapy in childhood diabetes mellitus. Guidelines for use. *Treat Endocrinol* 2003; 2: 11–21.
68. Boland E, Monsod T, Delucia M, et al. Limitations of conventional methods of self-monitoring of blood glucose: lessons learned from 3 days of continuous glucose sensing in pediatric patients with type 1 diabetes. *Diabetes Care* 2001; 24: 1858–1862.
69. Diabetes Research in Children Network Study Group. A randomized multicenter trial comparing the GlucoWatch biographer with standard glucose monitoring in children with type 1 diabetes. *Diabetes Care* 2005; 28: 1101–1106.
70. Sachedina N, Pickup J. Performance assessment of the Medtronic-Minimed continuous glucose monitoring system and its use for measurement of glycaemic control in type 1 diabetic subjects. *Diabet Med* 2003; 20: 1012–1015.
71. Diabetes Research in Children Network Study Group. Accuracy of the modified continuous glucose monitoring system sensor in an outpatient setting: results from a Diabetes Research in Children Network (DirecNet) study. *Diabetes Technol Ther* 2005; 7: 109–114.
72. Wilson D, Beck R, Tamborlane W, et al., and the DirecNet Study Group. The accuracy of the Freestyle Navigator continuous glucose monitoring system in children with type 1 diabetes. *Diabetes Care* 2007; 30: 59–64.
73. Ludvigsson J, Hanas R. Continuous subcutaneous glucose monitoring improved metabolic control in pediatric patients with type 1 diabetes: a controlled crossover study. *Pediatrics* 2003; 111: 933–938.
74. Diabetes Research in Children Network Study Group. Continuous glucose monitoring in children with type 1 diabetes. *J Pediatr* 2007; 151: 388–393.
75. Panteleon A, Loutseiko M, Steil G, Rebrin K. Evaluation of the effect of gain on meal response of an automated closed-loop insulin delivery system. *Diabetes* 2006; 55: 1995–2000.
76. Weinzimer S, Steil G, Swan K, et al. Fully automated closed loop insulin delivery vs. semi-automated hybrid control in pediatric patients with type 1 diabetes using an artificial pancreas. *Diabetes Care* 2008; 31: 934–939.

Toward a Cell-Based Cure for Diabetes: Advances in Production and Transplant of Beta Cells

Kathryn C. Claiborn and Doris A. Stoffers, MD, PhD

Institute for Diabetes, Obesity, and Metabolism, Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA

ABSTRACT

Type 1 diabetes results from autoimmune destruction of the insulin-producing beta cells of the pancreatic islets of Langerhans. Although developments in exogenous insulin therapy have greatly improved clinical outcomes in patients with diabetes, the ability of the pancreatic beta cell to exquisitely regulate the delivery of insulin and maintain normal levels of blood glucose is still far superior to what can be achieved by external delivery of insulin. As a result, the majority of patients with type 1 diabetes still experience the complications of chronic hyperglycemia or serious and potentially life-threatening hypoglycemia. The shortcomings of medical therapy have driven research toward more direct approaches of beta cell replacement. Indeed, the specificity of beta cell loss in type 1 diabetes makes this disease a particularly attractive candidate for cell-based therapies. In order for significant progress to be made, however, a thorough understanding of beta cell biology and more broadly islet biology is necessary. This review addresses recent advances in developmental biology that have expanded our understanding of islet cell differentiation, assesses the promise and limitations of islet transplantation, and discusses the future of alternative sources of beta cells, including directed differentiation of stem cells, replication of adult beta cells, and transdifferentiation of nonislet cells to a beta cell fate. ***Mt Sinai J Med 75:362–371, 2008.*** *© 2008 Mount Sinai School of Medicine*

Key Words: beta cell development, embryonic stem cell, islet transplantation, transdifferentiation.

Address Correspondence to:

Doris A. Stoffers
Institute for Diabetes, Obesity,
and Metabolism
Division of Endocrinology,
Diabetes, and Metabolism
Department of Medicine
University of Pennsylvania
School of Medicine
Philadelphia, PA
Email: stoffers@mail.
med.upenn.edu

A thorough understanding of pancreatic islet development is critical for the advancement of new cell-based therapies for diabetes. Indeed, knowledge derived from this work has already been instrumental in the development of strategies for directing embryonic stem (ES) cell differentiation toward a beta cell fate. The field has progressed dramatically in the last 10 years; in particular, the use of gene-targeting techniques in mice has rapidly expanded our knowledge of the genes involved in and required for normal pancreatic development.

Because of the similarity of the mouse pancreas to the human organ and the relative ease of genetic manipulation, much of the work that has defined pancreatic development has been performed in the mouse. In the mouse, pancreas development begins approximately on embryonic day 9 (E9) when dorsal and ventral patches of the gut epithelium evaginate.[1] These outgrowths are marked by expression of pancreatic and duodenal homeobox 1 (PDX1), a transcription factor required for subsequent pancreas development. Deletion of PDX1 results in the arrest of pancreatic development following budding, and a mutation in PDX1 has also been associated with pancreatic agenesis in humans.[2–4] The pancreatic buds undergo extensive growth and branching and eventually fuse into a single organ.[5]

The first endocrine cells of the pancreas express glucagon and appear coincident with the budding of the pancreatic anlagen, and 1 day later, insulin-positive cells are detectable.[5] Measurements of the replicative rate of these cells and lineage tracing

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20058

© 2008 Mount Sinai School of Medicine

experiments have suggested that adult beta and alpha cells do not arise from these early hormone-positive cells.[6] Expression analysis has revealed that glucagon messenger RNA is detectable as early as E8; pancreatic polypeptide, insulin, and ghrelin are detectable on E10.5, and somatostatin is detectable on E13.5.[7,8] Beginning on about E13.5, there is a dramatic increase in the level of insulin and exocrine enzymes in the pancreas (a phase of pancreatic development called the *secondary transition*), and it is at this time that insulin-positive cells up-regulate PDX1 and are seen to contain insulin granules, thus more closely resembling their mature counterparts.[5] These hormone-positive cells delaminate from the epithelium and form islets around E15.5.[9]

The basic-helix-loop-helix transcription factor Neurogenin 3 (NGN3) is considered a marker of all pancreatic endocrine cells, as illustrated by the phenotype of targeted disruption of the *Ngn3* gene in mice, in which no islets are present at birth.[10] NGN3 expression is controlled at least partly by hepatocyte nuclear factor 6 (HNF6); in vitro experiments suggest that HNF6 can bind to the NGN3 promoter and drive its expression. Mice with targeted disruption of the *Hnf6* gene have a dramatic reduction of NGN3-positive cells early in development and later have reduced numbers of islet cells and glucose intolerance.[11]

Many other transcription factors have been found to be required for the specification and survival of embryonic beta cells. In mice lacking paired box 4 (PAX4), very few insulin-positive and no somatostatin-positive cells are present at birth, but an increase in glucagon-positive cells has been observed.[12] The related factor PAX6 is also involved in endocrine cell development; *Pax6*$^{-/-}$ animals have a striking reduction in alpha cells, and *Pax4/Pax6* double-knockout animals express no endocrine hormones at all.[13] NK2 homeobox 2 (NKX2.2) is normally expressed in the dorsal bud early in pancreatic development and then restricted to endocrine cells. In *Nkx2.2*$^{-/-}$ mice, there is a dramatic reduction in islet cell mass, and insulin-positive cells are absent.[14] Expression analysis of the pancreatic epithelium from these mice revealed increased ghrelin expression, and immunohistochemistry showed that ghrelin-positive ϵ cells had replaced beta cells in this model. Likewise, *Pax4* null and *Pax6* mutant islets also have expanded populations of ghrelin-expressing cells.[8,15] Neurogenic differentiation 1 (NEUROD1), known to be a positive regulator of insulin transcription, is normally expressed in all islet cells of the pancreas. In *NeuroD1*$^{-/-}$ mice, there is a 75% reduction of beta cells at birth, and there is evidence of apoptosis throughout the islets, suggesting that NEUROD1 expression may contribute to beta cell survival.[16] Aristaless related homeobox (ARX) is a transcription factor first expressed on E9.5; *Arx*$^{-/-}$ mice display a complete loss of the alpha cell lineage and expansion of beta and somatostatin cells.[17]

Genetic mouse models have also shed light on the transcription factors required for maintenance of the beta cell phenotype. PDX1 expression, known to be required for pancreatic development, has also been shown to be required for maintenance of adult beta cells. In mice in which PDX1 expression is controlled by a tetracycline-sensitive transgene, overt diabetes develops within 2 weeks of PDX1 down-regulation concurrently with a decrease in islet size.[18] Similarly, Cre recombinase–mediated beta cell excision of PDX1 results in mice with severe hyperglycemia and beta cell loss.[19] Other factors found to be dispensable for the development of beta cells but required for adult function include V-maf musculoaponeurotic fibrosarcoma oncogene homolog A (MAFA), a transcription factor first identified biochemically as a component of the rat insulin promoter element 3b1 complex, which binds to and activates the insulin promoter.[20] Subsequent studies revealed that MAFA is required for mature beta cell function, and knockout animals develop normally but display markedly blunted insulin secretion. Myelin transcription factor 1 (MYT1) is a zinc finger protein expressed early in pancreatic development; deletion of the *Myt1* gene results in a normal number of embryonic endocrine cells, but these express multiple hormones. Similarly, a beta cell–specific loss of MYT1 results in adult islets in which many cells coexpress multiple hormones. In addition, the insulin-positive cells in this model are MAFB-positive, not MAFA-positive, and this suggests that MYT1 expression is required for maturation of several islet lineages.[21]

Significant evidence suggests that multiple signaling pathways are required for normal pancreas and islet development. In 1996, Gittes *et al.*[22] demonstrated that explants of pancreatic epithelium from E11 embryos could differentiate into acinar cells, ducts, and mature islets clusters when implanted under the kidney capsule with mesenchyme. Additional differentiation signals are likely received from the aorta, which is in close proximity to the pancreas primordium during development. In explant experiments, endoderm dissected on E8 expressed PDX1 and insulin when cocultured with an aorta.[23] Similarly, genetically manipulated *Xenopus* embryos that lack an aorta have reduced levels of PAX6, PDX1, and insulin.[23] Signals derived from the notochord are also required for pancreatic development; in cultured explants of dorsal pancreatic endoderm from chick embryos, removal of the adjacent notochord prevents the normal expression of PDX1 and PAX6.[24]

The previously described experiments demonstrate that mesenchymal and endodermal signals are required for normal islet development, but they do not elucidate the identity of those signals. These endothelial signals may be derived from the vascular basement membrane; cell sorting experiments have demonstrated that beta cells do not form their own basement membrane but rather depend on the collagen and laminin expressed by islet endothelial cells. Furthermore, Nikolova *et al.*[25] recently demonstrated that exposure to laminin increases insulin gene expression and proliferation in a beta cell line, and treatment with exogenous laminin partially restores insulin expression and proliferation rates in vascular endothelial growth factor A (VEGF-A) deficient islets, which have greatly reduced endothelial area.[26]

The hedgehog pathway is a cellular patterning signaling cascade with a complex and incompletely understood role in the promotion of an endocrine cell fate. Expression analysis of hedgehog signaling components in chick and mice has revealed that levels of the ligand Sonic hedgehog (SHH) and its receptor Patched are low or absent in the emerging pancreatic buds, although adjacent tissues show high levels of signaling.[27] The local repression of SHH is likely due to signals from the notochord; deletion of the notochord in chick embryos results in increased SHH expression in the pancreatic epithelium, and the resulting tissues do not develop into a mature pancreas.[24,27] Despite the requirement of repression early in pancreatic development, in zebrafish embryos exposed to hedgehog inhibitor cyclopamine, no pancreatic islet cells develop at all.[28] Thus, suppression of hedgehog signaling may be required in early specification of pancreatic progenitors, whereas up-regulation is required for commitment to endocrine cell fates.

Notochord-mediated suppression of hedgehog signaling may be the result of fibroblast growth factor (FGF) signaling molecules produced there. FGF signals in general are critical to pancreatic development and adult beta cell function. Hebrok *et al.*[27] demonstrated that culturing chick endoderm explants with exogenous FGF2 resulted in decreased expression of hedgehog signaling components and up-regulated PDX1 expression. The general requirement for FGF signals in pancreas development has also been demonstrated in zebrafish; morpholino knockdown of FGF3 and FGF8 in 2-cell embryos results in reduced insulin expression and disrupted islet architecture.[29] Furthermore, FGF signaling may contribute to control of retinoic acid (RA), a signaling molecule known to be involved in anteroposterior patterning in vertebrates. Treatment of zebrafish embryos with RA rescues insulin expression after FGF knockdown.[29] Pharmacological inhibition of RA signaling in *Xenopus* embryos inhibits the development of insulin-positive cells, and RA treatment of embryos increases the number of insulin-positive cells in a dose-dependent manner.[30] These endocrine cells appear at the expense of developing exocrine cells; RA likely works by inhibiting Notch signaling (thus promoting endocrine specification) and up-regulating PDX1.[30] In summary, precise control of multiple signaling pathways from surrounding tissues is required for proper development of the endocrine pancreas.

DIRECTION OF STEM CELL DIFFERENTIATION

ES cells are derived from blastocyst-stage embryos and can be maintained in culture while retaining the ability to differentiate into all cell types. Differentiation of ES cells into insulin-producing cells could feasibly provide an unlimited source of cells for cell-based diabetes therapy. Early work using mouse ES cells suggested that the production of insulin-producing cell clusters could be easily achieved by the adaptation of strategies of neural cell induction. Implantation of these cells prevented death in mice with streptozotocin (STZ)-induced diabetes, but the cells did not secrete enough insulin to rescue hyperglycemia.[31,32] These early successes were reconsidered, however, when Rajagopal and colleagues[33] demonstrated that the insulin staining observed ES cells was due to the uptake of exogenous insulin in the media.

More recent work in directing ES cell differentiation has primarily involved investigators striving to recapitulate the signaling and growth factor progression of normal pancreatic endocrine development in vitro.[34,35] In 2006, Xu *et al.*[34] demonstrated that human ES cells could be differentiated into PDX1-positive pancreatic progenitors by passing these cells through an embryoid body formation step. Later that year, D'Amour *et al.*[35] described a 5-step process that, using growth factors and various culture conditions, allowed progression from undifferentiated ES cells to definitive endoderm to hormone-expressing endocrine cells. This group used factors such as FGFs, hedgehog inhibitors, and RA to promote differentiation, and they assayed the development of endocrine cells by measuring the expression levels of transcription factors shown to be critical at each step in development, such as PDX1, NGN3, and PAX6. The insulin-positive cells thus generated had insulin content similar to that found in adult islet cells, as well as functional ATP-sensitive potassium channel and voltage-dependent calcium channels. Notable limitations of this achievement include the lack of glucose-responsive insulin secretion and the relatively low rate of transdifferentiation; insulin-positive cells made up only about 7% of the differentiated population.[35] Therefore, although researchers have already begun to capitalize on the findings of developmental biology to direct ES cell differentiation, much of the mystery of the specification and maturation of endocrine cells remains to be deduced by research in model organisms so that the process can be recapitulated in ES cells.

Work in human ES cells must always be considered with the caveats that their use and generation is restricted in the United States, and that the cell lines currently in use have been manipulated in culture for several years. The recent demonstration that human somatic cells can be reprogrammed to a pluripotent stem cell fate suggests that these ethical and legal issues may be circumvented.[36,37] Two groups of investigators used genes known to be enriched in ES cells and involved in the establishment of pluripotency to reprogram human somatic cells. Using slightly different protocols, both groups established pluripotent lines that were capable of differentiating into all 3 germ layers and forming teratomas when injected into mice. The limitations of this finding are considerable, given that the reprogramming involved the use of integrating lentiviral and retroviral vectors and in 1 case, the expression of the proto-oncogene c-Myc. However, if we assume that alternative methods of cell transduction can be developed, somatic cell–derived insulin-producing cells for diabetes therapy may be possible in the not too distant future.

TRANSDIFFERENTIATION

Another method of beta cell replacement for diabetes therapy might involve gene transfer or growth factor treatments to redirect nonislet cells to manufacture insulin and respond to changes in extracellular glucose. Given the close embryonic origins of the pancreas and liver, some investigators have focused efforts on transdifferentiation of the liver to a pancreatic endocrine cell fate. Building on the lessons of developmental biology, Ferber *et al.*[38] demonstrated that adenoviral-mediated expression of PDX1 induced insulin expression in the liver and rescued STZ-induced hyperglycemia in mice. Transgenic expression of PDX1 can induce endogenous PDX1 expression in the liver, thus promoting a prolonged effect of transiently expressing viral vectors.[39] Similarly, Kojima and colleagues[40] demonstrated that adenoviral delivery of the growth factor betacellulin and NeuroD, a transcription factor downstream of PDX1, rescued STZ-induced hyperglycemia and promoted the growth of islet like clusters within the liver.

This reprogramming of the liver into a pancreatic endocrine phenotype may involve a dedifferentiation step, as Meivar-Levy *et al.*[41] recently demonstrated that PDX1 overexpression results in repression of the transcription factor CCAAT/enhancer binding protein beta and up-regulation of immature hepatic factors such as alpha-fetoprotein. Likewise, in vitro experiments in adult rat hepatocytes have demonstrated that viral PDX1 expression induces insulin gene expression concurrently with a decrease in albumin levels.[42] Alternatively, these gene therapy approaches may be targeting the liver oval cells, which are considered to be resident liver stem cells. Yang *et al.*[43] demonstrated that liver oval cells could be isolated by fluorescence-activated cell sorting and differentiated into insulin-secreting cells by culturing in a high-glucose medium. In a small but compelling experiment, transplantation of these differentiated oval cells rescued hyperglycemia in an STZ-treated nonobese diabetic–severe combined immunodeficient mouse. Similarly, feeding mice a diet supplemented with the potent oval cell activator 3,5-diethoxycarbonyl-1,4-dihydrocollidine promoted recovery from STZ-induced hyperglycemia.[44] In this experiment, however, it is unclear whether differentiation of liver oval cells into insulin-producing cells actually accounts for the metabolic rescue.

Given their relative abundance and common origin from a PDX1-expressing pancreatic progenitor, several investigators have attempted to reprogram nonislet pancreatic epithelial cells to a beta cell fate. Transdifferentiation from a pancreatic acinar cell fate to a ductal cell fate has been well established in the rat model of ductal ligation, in which the ligated tail of the pancreas forms ductal complexes.[45,46] The finding of an expanded beta cell mass in this model suggests that acinar-to-islet transdifferentiation may also be occurring. Treatment with the hormone gastrin after duct ligation increases the beta cell mass in the transdifferentiating pancreas, and the ductal epithelium stains positively for the neuroendocrine marker protein gene product 9.5, which is normally expressed only in islets.[46] In 2005, Baeyens and colleagues[47] demonstrated that the treatment of rat pancreatic exocrine tissue with the growth factors epidermal growth factor (EGF) and leukemia inhibitory factor induced insulin expression in a fraction of cells. These insulin-positive cells were responsive to glucose and rescued alloxan-induced diabetes. Similarly, Bonner-Weir *et al.*[48] demonstrated that when human pancreatic ductal tissue is cultured on Matrigel, insulin-positive islet-like clusters emerge; these structures are capable of responding to changing glucose conditions by secreting insulin.

The aforementioned studies are suggestive of acinar- or ductal-to-islet transdifferentiation, but they do not formally prove the origin of newly formed beta cells. In experiments of cultured exocrine or ductal tissue, for example, the presence of contaminating beta cells that expand in response to growth factors could account for observed increases in the insulin transcript level. Minami *et al.*[49] demonstrated that when an acinar-enriched pancreatic fraction was cultured in the presence of nicotinamide and epidermal growth factor (EGF), an insulin-secreting population of cells emerged. Applying a lineage tracing system to the model, the group demonstrated that these insulin-positive cells were indeed derived from acinar cells that once expressed amylase and elastase. Lineage tracing has not yet demonstrated this transdifferentiation potential in vivo, however. Indeed, Desai *et al.*[50] recently demonstrated, using cre lox technology to mark acinar cells, that elastase-positive cells do not contribute to the generation of new beta cells after pancreatectomy. This finding suggests that in vivo the transdifferentiation capacity of adult acinar cells may be limited.

Although the reprogramming of liver or exocrine tissue to pancreatic endocrine cell fates rescues mouse models of hyperglycemia, the application to human disease may be limited by ongoing autoimmune attack on the beta cells of patients with type I diabetes face.[51] Insulin-producing cells in the liver might also be susceptible to autoimmune lymphocyte infiltration and apoptosis. Offering some optimism in this regard, Shternhall-Ron *et al.*[52] recently demonstrated that adenovirally delivered PDX1 rescued glucose tolerance in cyclophosphamide-accelerated nonobese diabetic mice. This finding suggests that gene therapy–mediated transdifferentiation may prove to be a viable strategy for producing new beta cells, despite ongoing autoimmune attack.

BETA CELL REGENERATION

Studies in mice and humans suggest that the beta cell mass, a critical component of glucose homeostasis, is plastic and can expand in response to increasing insulin demands. In humans, obese nondiabetic subjects have a 50% increase in beta cell volume in comparison with lean nondiabetic controls.[53] In a mouse model of genetically induced insulin resistance, the beta cell mass can expand up to 10-fold, thereby maintaining euglycemia.[54] Furthermore, in an autopsy series, Meier *et al.*[51] found insulin-producing beta cells in the pancreas of patients with longstanding type I diabetes, suggesting that even in the face of decades of autoimmune attack, the pancreas retained some regenerative capacity. These findings have led some researchers to speculate that a cure for diabetes may lie in the promotion of this innate beta cell regenerative activity.

The mechanism by which the beta cell mass is expanded has been hotly debated, although recent lineage tracing studies have provided compelling evidence that the adult beta cell mass is maintained by replication of existing beta cells rather than differentiation of stem or progenitor cells.[55–57] Dor *et al.*[55] developed an elegant transgenic lineage tracing model to genetically mark insulin-producing cells and identify their progeny. Using this system, they compared the percentage of labeled beta cells immediately after labeling to that after an extended chase period and found no dilution of labeled beta cells. Given the limited life span of the beta cells, these results suggested that the new beta cells were derived primarily by the replication of insulin-expressing beta cells. This group further demonstrated that beta cell replication could account for the regeneration of the beta cell mass following recovery from partial pancreatectomy. Nir *et al.*[58] recently employed the same lineage tracing system to demonstrate that beta cell replication mediated the recovery of the beta cell mass and glucose homeostasis after specific beta cell ablation. Teta *et al.*,[56] using the incorporation of the thymidine analogs 5-chloro-2'-deoxyuridine (CldU) and 5-iodo-2'doxyuridine (IdU) to mark replicating beta cells, found that after sequential 5-chloro-2'-deoxyuridine (CldU) and 5-iodo-2'doxyuridine (IdU) administration, the majority of cells were marked with one label or the other. These results suggested that new beta cells do not arise from a rapidly dividing precursor that would have incorporated both thymidine analogs.[56]

Capitalizing on the innate replicative ability of the beta cell would require an intimate understanding of the relevant stimuli and molecular players required for beta cell replication in vivo. Recent studies of the cell cycle mediators in beta cells have provided important clues to the regulation of beta cell replication and may suggest new therapeutic targets. The cell cycle regulator cyclin D2 appears to be a major player in postnatal beta cell replication; in *cyclin D2*$^{-/-}$ mice, beta cells do not replicate postnatally, and by 2 weeks after birth, animals have reduced beta cell mass and develop severe hyperglycemia.[59] Conversely, beta cell–directed overexpression of the cyclin-dependent kinase inhibitor p21 results in hyperglycemia and reduced beta cell proliferation, whereas inactivation of the p21 gene improves glycemic control in insulin-resistant *Irs2*$^{-/-}$ mice.[60] Thus, inhibition of p21 and promotion of cyclin D2 expression may prove to be viable strategies for the promotion of beta cell mass expansion.

In addition to cell cycle factors, transcription factors involved in the maintenance of adult beta cells may also represent targets for the promotion of beta cell mass expansion. In mice doubly heterozygous for insulin receptor and insulin receptor substrate 1 (*IR*+$^{/-}$ *Irs1*+$^{/-}$), the beta cell mass expands up to 10-fold in response to peripheral insulin resistance.[54] When *Pdx1* heterozygosity is imposed on this model, however, the beta cells fail to expand, and this suggests an important role for PDX1 in islet compensation for insulin resistance.[61] Similarly, imposing *Pdx1* heterozygosity on the *Glut4*+$^{/-}$ insulin-resistant mouse results in severely impaired glucose tolerance.[62] Although the expansion of the beta cell mass is similar in single and doubly heterozygous mice, *Glut4*$^{-/-}$ *Pdx1*+$^{/-}$ mice have decreased insulin production and impaired glucose-stimulated insulin release.[62] In the insulin-resistant *Irs2*$^{-/-}$ mouse model, haploinsuffiency of the forkhead transcription factor forkhead box O1 (*Foxo1*) improves glycemic control and results in an expansion of the beta cell mass. Gel shift and promoter assays suggest that Foxo1 may exert this effect by inhibiting PDX1 transcription.[63] Taken together, these studies suggest that maintenance of PDX1 levels are critical to an animal's ability to expand the beta cell mass and maintain beta cell function in the face of increasing insulin demand.

ISLET TRANSPLANTATION

The specificity of beta cell loss in diabetes makes this disease a particularly attractive candidate for cell replacement therapy. The first demonstration that an islet transplant could effectively ameliorate hyperglycemia occurred in 1972 when Ballinger and Lacy[64] implanted isolated islets intraperitoneally in STZ-induced diabetic rats. The group found that although recipient animals had improved hyperglycemia, the transplant failed to normalize blood sugar. Reasoning that islets normally release insulin into the portal circulation system, the same group of investigators demonstrated that islets infused intraportally could normalize blood sugar in STZ-treated rats.[65]

Significant advances in islet isolation techniques have allowed progress in research using transplanted human islets.[66] Many researchers have attempted portal vein islet infusions in conjunction with transplantation of other organs, particularly in conjunction with kidney transplants in patients with diabetic nephropathy. One such study evaluated the benefit of islet after kidney and simultaneous islet kidney transplants to type 1 diabetics.[67] Three months post-transplant, the majority of recipients had improved hemoglobin A1c levels and decreased fasting glucose. These patients remained insulin-dependent, but their daily insulin requirements had significantly decreased. Islet transplantation in conjunction with liver transplant was attempted in a study of patients who developed diabetes secondary to cirrhosis-induced insulin resistance.[68] In this experiment, the majority of islet recipients had reduced insulin requirements, and a few became insulin-independent.[68] These studies both suggested the promise of human islet transplantation therapy, but the absence of long-term follow-up of the patients makes it difficult to interpret their success.

Enthusiasm for the promise of human islet transplantation was ignited in 2000 with the publication of the Edmonton protocol. Because many traditional immunosuppression regimens were known to be toxic to islets, Shapiro *et al.*[69] employed a modified immunosuppression protocol that included daclizumab, a monoclonal antibody against interleukin 2, and eliminated glucocorticoids. In addition, study subjects received 2 to 3 islet infusions. In this first small study, all 7 participants achieved and maintained insulin independence for several months after transplantation.[69]

The excitement over this initial success was tempered, however, when follow-up studies suggested that euglycemia could not be maintained in the majority of patients. In a 5-year follow-up study of 65 patients who received islet transplants under the Edmonton protocol, only 10% remained insulin-independent.[70] Most patients did, however, have lasting improvements in glycemic control, with 80% maintaining islet graft function as measured by serum c-peptide levels.[70] The first international trial of this transplantation protocol demonstrated the technical difficulty of islet transplantation; success rates varied dramatically by location and correlated with the experience of the transplant and islet isolation teams.[71] Of the 36 patients included in this study, 44% were insulin-independent 1 year after transplantation, but only 14% remained so after 2 years. A small European study using the Edmonton protocol produced similar results, with only 30% of patients insulin-independent 1 year after transplant.[72] The authors of this study, however, emphasize that insulin independence may not be the only outcome that should be considered after islet transplant; 80% of patients evaluated had a "good metabolic outcome," which was defined as decreased insulin requirements and fewer hypoglycemic events 2 years after transplant.

Some of the most challenging complications of islet transplantation arise from the need for ongoing immunosuppression. Side effects of immunosuppressant drugs include mouth ulcers, decreased renal function, and hypertension.[73] In addition, the immunosuppressive agents used in the Edmonton protocol are known to decrease the proliferative capacity of beta cells.[58] One option for avoiding this complication is to deliver encapsulated islets; immobilized in selectively permeable casings, islets would be capable of sensing glucose and secreting insulin but would be protected from immune attack. In theory, these membranes would permit islet transplantation without immunosuppression and would also make xenotransplantation a viable option, potentially solving the problem of donor scarcity by making use of more abundant sources such as porcine islets. Microencapsulation strategies have been used successfully in animal studies; the delivery of bovine islets packaged in alginate gel spheres to the peritoneum of rats has been shown to normalize STZ-induced hyperglycemia.[74] Although complications with vascularization of and fibrosis around encapsulated islets have not been totally resolved, clinical trials for encapsulated islet delivery to type 1 diabetics are moving forward.[75]

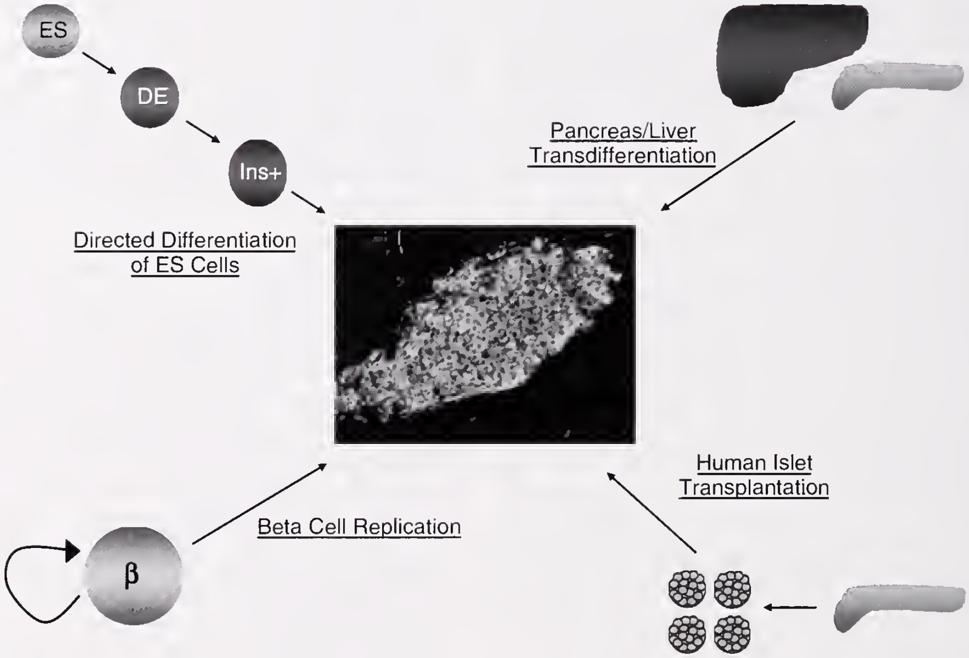

ES
DE
Ins+
Pancreas/Liver Transdifferentiation
Directed Differentiation of ES Cells
Human Islet Transplantation
Beta Cell Replication
β

Fig 1. Alternative sources of new beta cells. Type 1 diabetes results from the autoimmune destruction of pancreatic beta cells. A cure for the disease may lie in the development of techniques to replenish islet mass. The application of lessons from developmental biology has already permitted researchers to direct the differentiation of ES cells through a DE step toward an endocrine (Ins+) cell fate. The potential exists for liver and nonendocrine pancreatic cells to be reprogrammed by gene transfer or growth factor treatments to produce and release insulin in response to glucose. Recent evidence has suggested that adult beta cells are mitotically active and that replication of these cells can occur in response to increasing insulin demands. A complete understanding of the factors and signals that control this process could allow this innate ability to be harnessed in patients with diabetes. Finally, cadaveric human islets represent a promising source of therapeutic beta cells through transplantation. Immunofluorescence staining of a primary adult mouse islet is shown in the center (insulin appears green, and endothelial marker Meca-32 appears red). **Abbreviations:** DE, definitive endoderm; ES, embryonic stem; Ins+, insulin-positive. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.].

CONCLUDING REMARKS

The specific destruction of beta cells that characterizes type 1 diabetes suggests that a cure for the disease simply requires replacement of those cells. Recent breakthroughs in islet transplantation demonstrate the power of this cell replacement to re-establish glucose homeostasis, but sources of these islets are extremely limited, and thus far, the vast majority of islet recipients have experienced a failure to maintain insulin independence, despite multiple islet infusions. Thus, the development of alternative sources of islets, such as the promotion of transdifferentiation of non-beta cells into an insulin-secreting phenotype, may be viable option (Figure 1). In addition, advances in the understanding of pancreas development have already informed the development of techniques to direct the differentiation of ES cells toward endocrine cell fates. The pace of discovery in the last decade has been striking, but additional progress in understanding the genes and signals that direct beta cell development, as well as true tests of the replicative potential of the adult beta cell, will be required to achieve a sustainable, cell-based therapy for diabetes.

ACKNOWLEDGMENT

The authors acknowledge support from the National Institutes of Health (R01 DK068157, R01 DK06296502, and P01 DK49210).

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Murtaugh LC, Melton DA. Genes, signals, and lineages in pancreas development. *Annu Rev Cell Dev Biol* 2003; 19: 71–89.
2. Jonsson J, Carlsson L, Edlund T, Edlund H. Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature* 1994; 371: 606–609.
3. Offield MF, Jetton TL, Labosky PA, et al. PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development* 1996; 122: 983–995.
4. Stoffers DA, Zinkin NT, Stanojevic V, et al. Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence. *Nat Genet* 1997; 15: 106–110.
5. Pictet R, Rutter WJ. Development of the embryonic endocrine pancreas. In: Steiner NFD, ed. *Handbook of Physiology*. Baltimore, MD: Williams & Williams; 1972:; 25–66.
6. Herrera PL. Adult insulin- and glucagon-producing cells differentiate from two independent cell lineages. *Development* 2000; 127: 2317–2322.
7. Herrera PL, Huarte J, Sanvito F, et al. Embryogenesis of the murine endocrine pancreas; early expression of pancreatic polypeptide gene. *Development* 1991; 113: 1257–1265.
8. Heller RS, Jenny M, Collombat P, et al. Genetic determinants of pancreatic epsilon-cell development. *Dev Biol* 2005; 286: 217–224.
9. Kaung HL. Growth dynamics of pancreatic islet cell populations during fetal and neonatal development of the rat. *Dev Dyn* 1994; 200: 163–175.
10. Gradwohl G, Dierich A, LeMeur M, Guillemot F. Neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. *Proc Natl Acad Sci U S A* 2000; 97: 1607–1611.
11. Jacquemin P, Durviaux SM, Jensen J, et al. Transcription factor hepatocyte nuclear factor 6 regulates pancreatic endocrine cell differentiation and controls expression of the proendocrine gene ngn3. *Mol Cell Biol* 2000; 20: 4445–4454.
12. Sosa-Pineda B, Chowdhury K, Torres M, et al. The Pax4 gene is essential for differentiation of insulin-producing beta cells in the mammalian pancreas. *Nature* 1997; 386: 399–402.
13. St-Onge L, Sosa-Pineda B, Chowdhury K, et al. Pax6 is required for differentiation of glucagon-producing alpha-cells in mouse pancreas. *Nature* 1997; 387: 406–409.
14. Sussel L, Kalamaras J, Hartigan-O'Connor DJ, et al. Mice lacking the homeodomain transcription factor Nkx2.2 have diabetes due to arrested differentiation of pancreatic beta cells. *Development* 1998; 125: 2213–2221.
15. Prado CL, Pugh-Bernard AE, Elghazi L, et al. Ghrelin cells replace insulin-producing beta cells in two mouse models of pancreas development. *Proc Natl Acad Sci U S A* 2004; 101: 2924–2929.
16. Naya FJ, Huang HP, Qiu Y, et al. Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-deficient mice. *Genes Dev* 1997; 11: 2323–2334.
17. Collombat P, Mansouri A, Hecksher-Sorensen J, et al. Opposing actions of Arx and Pax4 in endocrine pancreas development. *Genes Dev* 2003; 17: 2591–2603.
18. Holland AM, Góñez LJ, Naselli G, et al. Conditional expression demonstrates the role of the homeodomain transcription factor Pdx1 in maintenance and regeneration of beta-cells in the adult pancreas. *Diabetes* 2005; 54: 2586–2595.
19. Ahlgren U, Jonsson J, Jonsson L, et al. Beta-cell-specific inactivation of the mouse Ipf1/Pdx1 gene results in loss of the beta-cell phenotype and maturity onset diabetes. *Genes Dev* 1998; 12: 1763–1768.
20. Matsuoka TA, Zhao L, Artner I, et al. Members of the large Maf transcription family regulate insulin gene transcription in islet beta cells. *Mol Cell Biol* 2003; 23: 6049–6062.
21. Wang S, Zhang J, Zhao A, et al. Loss of Myt1 function partially compromises endocrine islet cell differentiation and pancreatic physiological function in the mouse. *Mech Dev* 2007; 124: 898–910.
22. Gittes GK, Galante PE, Hanahan D, et al. Lineage-specific morphogenesis in the developing pancreas:

role of mesenchymal factors. *Development* 1996; 122: 439–447.

23. Lammert E, Cleaver O, Melton D. Induction of pancreatic differentiation by signals from blood vessels. *Science* 2001; 294: 564–567.
24. Kim SK, Hebrok M, Melton DA. Notochord to endoderm signaling is required for pancreas development. *Development* 1997; 124: 4243–4252.
25. Nikolova G, Jabs N, Konstantinova I, et al. The vascular basement membrane: a niche for insulin gene expression and beta cell proliferation. *Dev Cell* 2006; 10: 397–405.
26. Lammert E, Gu G, McLaughlin M, et al. Role of VEGF-A in vascularization of pancreatic islets. *Curr Biol* 2003; 13: 1070–1074.
27. Hebrok M, Kim SK, Melton DA. Notochord repression of endodermal Sonic hedgehog permits pancreas development. *Genes Dev* 1998; 12: 1705–1713.
28. diIorio PJ, Moss JB, Sbrogna JL, et al. Sonic hedgehog is required early in pancreatic islet development. *Dev Biol* 2002; 244: 75–84.
29. Song J, Kim HJ, Gong Z, et al. Vhnf1 acts downstream of Bmp, Fgf, and RA signals to regulate endocrine beta cell development in zebrafish. *Dev Biol* 2007; 303: 561–575.
30. Chen Y, Pan FC, Brandes N, et al. Retinoic acid signaling is essential for pancreas development and promotes endocrine at the expense of exocrine cell differentiation in Xenopus. *Dev Biol* 2004; 271: 144–160.
31. Lumelsky N, Blondel O, Laeng P, et al. Differentiation of embryonic stem cells to insulin-secreting structures similar to pancreatic islets. *Science* 2001; 292: 1389–1394.
32. Hori Y, Rulifson IC, Tsai BC, et al. Growth inhibitors promote differentiation of insulin-producing tissue from embryonic stem cells. *Proc Natl Acad Sci U S A* 2002; 99: 16105–16110.
33. Rajagopal J, Anderson WJ, Kume S, et al. Insulin staining of ES cell progeny from insulin uptake. *Science* 2003; 299: 363.
34. Xu X, Kahan B, Forgianni A, et al. Endoderm and pancreatic islet lineage differentiation from human embryonic stem cells. *Cloning Stem Cells* 2006; 8: 96–107.
35. D'Amour KA, Bang AG, Eliazer S, et al. Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. *Nat Biotechnol* 2006; 24: 1392–1401.
36. Yu J, Vodyanik MA, Smuga-Otto K, et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science* 2007; 318: 1917–1920.
37. Takahashi K, Tanabe K, Ohnuki M, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 2007; 131: 861–872.
38. Ferber S, Halkin A, Cohen H, et al. Pancreatic and duodenal homeobox gene 1 induces expression of insulin genes in liver and ameliorates streptozotocin-induced hyperglycemia. *Nat Med* 2000; 6: 568–572.
39. Ber I, Shternhall K, Perl S, et al. Functional, persistent, and extended liver to pancreas transdifferentiation. *J Biol Chem* 2003; 278: 31950–31957.
40. Kojima H, Fujimiya M, Matsumura K, et al. NeuroD-betacellulin gene therapy induces islet neogenesis in the liver and reverses diabetes in mice. *Nat Med* 2003; 9: 596–603.
41. Meivar-Levy I, Sapir T, Gefen-Halevi S, et al. Pancreatic and duodenal homeobox gene 1 induces hepatic dedifferentiation by suppressing the expression of CCAAT/enhancer-binding protein beta. *Hepatology* 2007; 46: 898–905.
42. Yamada S, Yamamoto Y, Nagasawa M, et al. In vitro transdifferentiation of mature hepatocytes into insulin-producing cells. *Endocr J* 2006; 53: 789–795.
43. Yang L, Li S, Hatch H, et al. In vitro trans-differentiation of adult hepatic stem cells into pancreatic endocrine hormone-producing cells. *Proc Natl Acad Sci U S A* 2002; 99: 8078–8083.
44. Kim S, Shin JS, Kim HJ, et al. Streptozotocin-induced diabetes can be reversed by hepatic oval cell activation through hepatic transdifferentiation and pancreatic islet regeneration. *Lab Invest* 2007; 87: 702–712.
45. Wang RN, Kloppel G, Bouwens L. Duct- to islet-cell differentiation and islet growth in the pancreas of duct-ligated adult rats. *Diabetologia* 1995; 38: 1405–1411.
46. Rooman I, Lardon J, Bouwens L. Gastrin stimulates beta-cell neogenesis and increases islet mass from transdifferentiated but not from normal exocrine pancreas tissue. *Diabetes* 2002; 51: 686–690.
47. Baeyens L, De Breuck S, Lardon J, et al. In vitro generation of insulin-producing beta cells from adult exocrine pancreatic cells. *Diabetologia* 2005; 48: 49–57.
48. Bonner-Weir S, Taneja M, Weir GC, et al. In vitro cultivation of human islets from expanded ductal tissue. *Proc Natl Acad Sci U S A* 2000; 97: 7999–8004.
49. Minami K, Okuno M, Miyawaki K, et al. Lineage tracing and characterization of insulin-secreting cells generated from adult pancreatic acinar cells. *Proc Natl Acad Sci U S A* 2005; 102: 15116–15121.
50. Desai BM, Oliver-Krasinski J, De Leon DD, et al. Preexisting pancreatic acinar cells contribute to acinar cell, but not islet beta cell, regeneration. *J Clin Invest* 2007; 117: 971–977.
51. Meier JJ, Bhushan A, Butler AE, et al. Sustained beta cell apoptosis in patients with long-standing type 1 diabetes: indirect evidence for islet regeneration? *Diabetologia* 2005; 48: 2221–2228.
52. Shternhall-Ron K, Quintana FJ, Perl S, et al. Ectopic PDX-1 expression in liver ameliorates type 1 diabetes. *J Autoimmun* 2007; 28: 134–142.
53. Butler AE, Janson J, Bonner-Weir S, et al. Beta-cell deficit and increased beta-cell apoptosis in humans with type 2 diabetes. *Diabetes* 2003; 52: 102–110.
54. Brüning JC, Winnay J, Bonner-Weir S, et al. Development of a novel polygenic model of NIDDM in mice heterozygous for IR and IRS-1 null alleles. *Cell* 1997; 88: 561–572.
55. Dor Y, Brown J, Martinez OI, Melton DA. Adult pancreatic beta-cells are formed by self-duplication rather than stem-cell differentiation. *Nature* 2004; 429: 41–46.
56. Teta M, Rankin MM, Long SY, et al. Growth and regeneration of adult beta cells does not involve specialized progenitors. *Dev Cell* 2007; 12: 817–826.
57. Brennand K, Huangfu D, Melton D. All beta cells contribute equally to islet growth and maintenance. *PLoS Biol* 2007; 5: e163.
58. Nir T, Melton DA, Dor Y. Recovery from diabetes in mice by beta cell regeneration. *J Clin Invest* 2007; 117: 2553–2561.

59. Georgia S, Bhushan A. Beta cell replication is the primary mechanism for maintaining postnatal beta cell mass. *J Clin Invest* 2004; 114: 963–968.
60. Uchida T, Nakamura T, Hashimoto N, et al. Deletion of Cdkn1b ameliorates hyperglycemia by maintaining compensatory hyperinsulinemia in diabetic mice. *Nat Med* 2005; 11: 175–182.
61. Kulkarni RN, Jhala US, Winnay JN, et al. PDX-1 haploinsufficiency limits the compensatory islet hyperplasia that occurs in response to insulin resistance. *J Clin Invest* 2004; 114: 828–836.
62. Brissova M, Blaha M, Spear C, et al. Reduced PDX-1 expression impairs islet response to insulin resistance and worsens glucose homeostasis. *Am J Physiol Endocrinol Metab* 2005; 288: E707–E714.
63. Kitamura T, Nakae J, Kitamura Y, et al. The forkhead transcription factor Foxo1 links insulin signaling to Pdx1 regulation of pancreatic beta cell growth. *J Clin Invest* 2002; 110: 1839–1847.
64. Ballinger WF, Lacy PE. Transplantation of intact pancreatic islets in rats. *Surgery* 1972; 72: 175–186.
65. Kemp CB, Knight MJ, Scharp DW, et al. Transplantation of isolated pancreatic islets into the portal vein of diabetic rats. *Nature* 1973; 244: 447.
66. Ricordi C, Lacy PE, Finke EH, et al. Automated method for isolation of human pancreatic islets. *Diabetes* 1988; 37: 413–420.
67. Bretzel RG, Brandhorst D, Brandhorst H, et al. Improved survival of intraportal pancreatic islet cell allografts in patients with type-1 diabetes mellitus by refined peritransplant management. *J Mol Med* 1999; 77: 140–143.
68. Angelico MC, Alejandro R, Nery J, et al. Transplantation of islets of Langerhans in patients with insulin-requiring diabetes mellitus undergoing orthotopic liver transplantation–the Miami experience. *J Mol Med* 1999; 77: 144–147.
69. Shapiro AM, Lakey JR, Ryan EA, et al. Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen. *N Engl J Med* 2000; 343: 230–238.
70. Ryan EA, Paty BW, Senior PA, et al. Five-year follow-up after clinical islet transplantation. *Diabetes* 2005; 54: 2060–2069.
71. Shapiro AM, Ricordi C, Hering BJ, et al. International trial of the Edmonton protocol for islet transplantation. *N Engl J Med* 2006; 355: 1318–1330.
72. Badet L, Benhamou PY, Wojtusciszyn A, et al. Expectations and strategies regarding islet transplantation: metabolic data from the GRAGIL 2 trial. *Transplantation* 2007; 84: 89–96.
73. Rother KI, Harlan DM. Challenges facing islet transplantation for the treatment of type 1 diabetes mellitus. *J Clin Invest* 2004; 114: 877–883.
74. Lanza RP, Ecker DM, Kühtreiber WM, et al. Transplantation of islets using microencapsulation: studies in diabetic rodents and dogs. *J Mol Med* 1999; 77: 206–210.
75. Calafiore R, Basta G, Luca G, et al. Standard technical procedures for microencapsulation of human islets for graft into nonimmunosuppressed patients with type 1 diabetes mellitus. *Transplant Proc* 2006; 38: 1156–1157.

Kidney and Pancreas Transplantation in Type 1 Diabetes Mellitus

Susan M. Lerner, MD

Recanati/Miller Transplant Institute, Mount Sinai School of Medicine, New York, NY

ABSTRACT

Type 1 diabetes mellitus is characterized by autoimmune destruction of pancreatic beta cells, leading to a state of absolute insulin deficiency. Glycemic control via the use of exogenous insulin injections is often imperfect, resulting in multiple long-term complications, such as retinopathy, neuropathy, vasculopathy, and nephropathy. The Diabetes Control and Complications Trial has provided conclusive evidence that better glycemic control by intensive insulin treatment effectively delays the onset and slows the progression of diabetic retinopathy, nephropathy, and neuropathy in patients with insulin-dependent diabetes mellitus. At this moment, the only reliable option for achieving long-term insulin independence is whole-pancreas transplantation. The proposed benefits of pancreas transplantation are clear: improved quality of life, prevention of recurrent diabetic nephropathy, freedom from exogenous insulin with euglycemia and normalization of glycosylated hemoglobin, less stringent dietary restrictions, less frequent blood glucose monitoring, and stabilization of or improvement in secondary complications. The trade-offs to the patient are the operative risk, the need for chronic immunosuppression, and the inherent side effects of chronic immunosuppression. ***Mt Sinai J Med 75:372–384, 2008.*** © *2008 Mount Sinai School of Medicine*

Key Words: complications of transplantation, diabetes mellitus, glycemic control, immunosuppression, islet transplantation, pancreas transplantation.

Address Correspondence to:

Susan M. Lerner
Recanati/Miller Transplant Institute
Mount Sinai Medical Center
New York, NY
Email: susan.lerner@mountsinai.org

Diabetes mellitus type 1 is characterized by autoimmune destruction of pancreatic beta cells, leading to a state of absolute insulin deficiency. According to the World Health Organization, approximately 1.4 million individuals are affected with type 1 diabetes mellitus in the United States.[1] The isolation and development of insulin as a treatment for diabetes changed the disease from one that is rapidly fatal to a chronic disease with the potential for multiple secondary complications. Glycemic control via the use of exogenous insulin injections is imperfect, resulting in long-term complications such as retinopathy, neuropathy, vasculopathy, and, in 20% to 30% of cases, diabetic nephropathy.[2] Chronically abnormal lipid metabolism, accelerated atherosclerosis, and destruction of the microvascular system result in global vascular disease, which leads to amputations and premature death from myocardial infarctions and cerebrovascular accidents.[3]

The earliest clinical manifestation of diabetic renal disease is microalbuminuria. Without specific interventions, within 10 to 15 years, about 80% of type 1 diabetic patients with sustained microalbuminuria progress to overt nephropathy. Approximately 30% of patients with diabetes and end-stage renal disease who enter renal replacement therapy have significant coronary artery disease. Mortality rates are high in patients with type 1 diabetes mellitus. The mortality rate is 10% to 13% after 20 years of disease with a 6% rate of renal replacement therapy.[4] During the first 20 years of the disease, most of the excess mortality is attributed to renal failure; after this period, it is considered to result largely from cardiovascular and cerebrovascular events. Factors associated with a greater risk of mortality are higher age, higher hemoglobin A1c values, presence of microalbuminuria or macroalbuminuria, lower creatinine clearance, and male gender.[4] Hemoglobin A1c, or glycated hemoglobin, is a form of hemoglobin used primarily to identify the average plasma glucose concentration over prolonged periods of time.

The Diabetes Control and Complications Trial provided conclusive evidence that better glycemic

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20056

© 2008 Mount Sinai School of Medicine

control by intensive insulin treatment effectively delays the onset and slows the progression of diabetic retinopathy, nephropathy, and neuropathy in patients with type 1 diabetes.[1] Prevention of microvascular and macrovascular diabetic complications can be obtained by optimal glycemic control. At this moment, the only reliable option for achieving long-term insulin independence is whole-pancreas transplantation.

The world's first clinical pancreas transplant was performed, simultaneously with a kidney graft, in 1966 at the University of Minnesota to treat a diabetic uremic patient. Although the early results were poor, this first transplant did demonstrate that glucose control without exogenous insulin was possible. The procedure established that endogenous secretion of insulin with normal feedback mechanisms could occur with a pancreas transplant.[2] The frequency of pancreas transplantation has increased as the results have improved, so that over 25,000 pancreas transplants have now been performed, with almost 1400 pancreas transplants performed in 2006 in the United States.[5]

TYPES AND TIMING OF TRANSPLANTATION

Pancreas transplants are performed in 3 different categories of diabetic recipients, who are classified according to their kidney function status. The majority of pancreas transplants have been and continue to be performed as simultaneous pancreas-kidney (SPK) transplants for patients with type 1 diabetes mellitus and end-stage renal disease. In recent years, there has been a relative increase in pancreas after kidney (PAK) transplantation and pancreas transplant alone (PTA). A PAK transplant involves transplantation of a pancreatic graft into a recipient with a well-functioning living or cadaveric kidney allograft. PTA is generally reserved for patients in whom glucose control is extremely difficult to achieve and who suffer from frequent episodes of life-threatening hypoglycemia. The Centers for Medicare and Medicaid Services issued a national coverage determination that stated that Medicare would pay for PTA. This decision allowed private insurers to follow suit and liberalize the performance of this procedure.[6] The proposed benefits of pancreas transplantation are clear: improved quality of life, prevention of recurrent diabetic nephropathy, freedom from exogenous insulin with euglycemia and normalization of glycosylated hemoglobin, less stringent dietary restrictions, less frequent blood glucose monitoring, and stabilization of or improvement in secondary complications. No other regimen of insulin delivery can achieve this level of physiologic regulation. The trade-offs to the patient are the operative risk, the need for chronic immunosuppression, and the inherent side effects of chronic immunosuppression.

Pancreas transplant recipient and graft survival rates have improved in all 3 categories since the introduction of the procedure more than 3 decades ago, and this has led to an increase in the number of candidates and an increase in the number of transplants performed in all 3 categories. Current, unadjusted pancreas graft survival for SPK transplants is 86% and 71% at 1 and 5 years, respectively. One- and five-year survival rates for solitary grafts continue to improve and currently are 78% and 57% for PAK transplants and 77% and 56% for PTA transplants. Patient survival is worse compared to that of waitlisted patients in the first 90 days after transplantation, reflecting the stress of surgery; however, at 4 years after transplantation, survival for SPK recipients far exceeds that of their waitlisted counterparts and is no worse for PAK and PTA recipients. The most influential factors leading to patient mortality have been shown to be graft failure and return to exogenous insulin. Thus, to reduce the posttransplant mortality, preventing immunologic and nonimmunologic causes of graft failure is paramount.[7] The leading causes of death are cardiovascular and cerebrovascular disease. In addition, in comparison with type 1 diabetic recipients of a kidney alone, there appears to be an increase in infectious deaths.

Pancreas transplants today are most common for diabetic patients who have the secondary complication of kidney failure. Options for such patients include receiving both organs simultaneously or sequentially. SPK transplants have rendered recipients free of both insulin and dialysis with a single operation. However, PAK transplants have several advantages over SPK transplants. First, one significant advantage is the ability to use a living donor for the kidney transplant. Several studies have documented the advantage of a living donor (versus a cadaver) in all renal transplants. The incidence of kidney graft loss due to chronic rejection has also been shown to be higher for SPK transplants, again likely because of the use of cadaver donors. Graft loss due to technical complications is also higher in SPK recipients. Using living donors also expands the donor pool for kidneys, and this benefits patients on both the cadaver SPK and kidney alone waiting lists.[8]

Using living donors also allows for a preemptive transplant (before the patient needs to be on dialysis), Studies have shown that recipients who had less than 6 months of dialysis before a renal transplant had much better 5- and 10-year unadjusted allograft survival rates than those who spent more than 2 years on dialysis. In other studies, the waiting time on dialysis was the strongest independent modifiable risk factor for renal transplant outcomes. The most plausible explanation for this advantage is the avoidance of dialysis-associated morbidities with preemptive transplantation. In addition, the increasing duration of dialysis was associated with increasing odds of rejection.[9] Preemptive SPK transplantation has also been observed to be associated with improved kidney allograft survival in comparison with nonpreemptive SPK transplantation. Transplantation prior to the initiation of chronic dialysis is associated with a lower rate of kidney allograft failure among SPK recipients.[10]

Another advantage is that PAK transplants seem to be better tolerated by most recipients. A shorter operation and less extensive dissection may play a role. In addition, PAK recipients may also be in better condition at the time of their transplant, having the advantages of a functioning kidney and of being nonuremic at the time of transplant. Finally, the waiting time for PAK transplants is significantly lower. The demand is much higher in the United States for cadaver kidneys, and therefore it is the need for a cadaver kidney in an SPK transplant that determines the waiting time.[8]

Another question is then that of optimal timing of a PAK transplant. There does not seem to be any difference in patient or graft survival rates or in surgical complications whether the pancreas transplant is performed early (less than 4 months) or late (more than 4 months) after the kidney transplant. The current recommendation is to perform the pancreas transplant as soon as the recipient adequately recovers from the kidney transplant operation and has stable renal function.[8] Another option is to perform the living donor kidney transplant at the same time as the cadaver pancreas transplant. The main drawback to simultaneous pancreas–living kidney transplantation is the coordination of a living donor nephrectomy with a cadaver pancreas transplant, which requires surgical expertise for this procedure 24 hours a day.[11]

The main disadvantage of PAK transplants is that the long-term pancreas graft survival rates remain slightly lower than those of SPK transplants. Pancreas graft survival rates were not found to be significantly different at 1 year between the 2 groups, but they decreased by 3 years post-transplant. This difference is likely due to immunologic causes: the incidence of pancreas graft loss due to acute and chronic rejection was found to be significantly higher for PAK recipients. It may be that rejection episodes are missed or diagnosed later because of the inability to use serum creatinine levels as a marker for rejection (because the kidney and pancreas grafts are from different donors).[8] Some studies have suggested that the immunologic problems which may have limited PAK transplantation may be ameliorated with newer immunosuppressive regimens, and this could lead to a reduction in the rejection rates.[12]

Of the more than 15,000 transplants performed from 1987 to 2004, 5% were retransplants. Most of the PAK retransplants (71%) have been in recipients of SPK transplants whose primary pancreas graft failed while the kidney continued to function. The proportion of PTA cases that are retransplants has slightly declined, and this reflects the improvement in the results of primary pancreas transplants. Of the pancreas retransplants in the SPK category, 68% had a prior SPK transplant, 14% had a prior PAK transplant, and the remainder had a prior PTA (18%).[13]

IMPACT OF PANCREAS TRANSPLANTATION

Successful pancreas transplantation normalizes glycemic control and may have beneficial effects on diabetic macrovascular complications. Epidemiologic studies using registry data have shown that SPK transplantation confers survival advantage over diseased donor kidney transplantation alone and that the death rate caused by cardiovascular disease is significantly decreased in SPK patients. Pancreas transplants may have an anti-inflammatory role that results in improved endothelial function and decreased arterial stiffness.[14] Studies have shown that the progression of coronary atherosclerosis in patients with a functioning pancreas graft is reduced in comparison with patients in whom the pancreas graft has failed. Comparable changes have been identified in carotid arteries as well. The pathogenesis of atherosclerosis is complex. Hyperglycemia is believed to contribute to atherosclerotic plaque formation in a number of ways, including glycation of collagen, alteration of endothelial cell function, glycoxidation of low-density lipoprotein cholesterol, and increased platelet reactivity. Therefore, pancreas transplantation may, apart from slowing down the progression of coronary atherosclerosis, also stabilize atherosclerotic plaques and restore endothelial reactivity, thereby reducing clinical atherosclerotic events as well.[15] In addition, in an examination of atherosclerotic risk factors several years after transplantation, kidney/pancreas recipients were found to have decreased levels of von Willebrand's factor, fibrinogen, triglyceride, D-dimers, and homocysteine in comparison with recipients of kidney transplant alone (KTA).[2]

The effect of pancreas transplantation on pulmonary function has been explored as well. Reduced lung volumes and impaired diffusing capacity have been demonstrated in type 1 diabetic patients. The cause of this pulmonary dysfunction is still widely speculative. Both alterations of connective tissue components such as collagen and elastin and microangiopathy of pulmonary vessels have been proposed. A clear improvement in pulmonary function has been demonstrated after SPK transplantation, but it has not been determined whether the elimination of diabetes or uremia contributes more to this improvement.[16]

Diabetic nephropathy is the most common single cause of end-stage renal disease in the United States. The earliest clinical sign of nephropathy is microalbuminuria. Without specific interventions, 80% of subjects with type 1 diabetes who develop sustained microalbuminuria have their urinary albumin excretion increase over a period of 10 to 15 years, with hypertension also developing along the way. Eventually, chronic kidney disease develops in 50% of type 1 diabetic individuals within 10 years and in greater than 75% by 20 years. In addition to its being the earliest manifestation of nephropathy, proteinuria is a marker of greatly increased cardiovascular morbidity and mortality for patients with diabetes.[17] Pancreas transplantation has a positive impact on diabetic nephropathy. By 2 years post-transplant, KTA recipients show elevated creatinine and urine albumin levels compared with those in SPK recipients. Furthermore, on biopsy, the glomerular basement membrane changes that are seen in KTA recipients are not present in SPK recipients.[18] This suggests that prolonged normoglycemia prevents the recurrence of diabetic glomerulopathy in kidney allografts.[2] Restoration of endogenous insulin, the consequent normalization of blood glucose levels, and improvement of other metabolic parameters all play fundamental roles in this improved renal function. Improvement of glucose levels has been associated with reduced hyperfiltration and a diminished albumin excretion rate.[17] Therefore, successful PTA determines the disappearance of nephrotic syndrome, as documented by the disappearance or significant improvement of the clinical signs of this condition.[19]

Diabetic retinopathy is a leading cause of adult blindness. The severity of diabetic retinopathy is strongly associated with the duration of diabetes, hypertension, renal disease, and blood glucose control. Five years after diagnosis, 23% of diabetic patients have retinopathy.[20] Diabetic retinopathy is a microangiopathy affecting the precapillary arterioles, capillaries, and venules, causing microvascular occlusion and leakage. Microaneurysm formation due to damaged supporting capillary pericytes and localized dilatation is an early feature of diabetic retinopathy, which is followed by capillary closure, retinal ischemia, retinal edema, hemorrhage, and hard exudate formation secondary to vascular leakage. In more advanced retinopathy, photocoagulation is effective in stabilizing vision, but reversal of disease with normoglycemia in these patients is unlikely because of irreversible injury.[21] The results of studies of diabetic retinopathy after pancreas transplantation are difficult to interpret, being confounded by small patient numbers, short follow-up, and differing control groups.[22] Successful pancreas transplantation can positively affect the course of diabetic retinopathy, mostly through stabilization of the disease.[23] Although stabilization has been achieved by some patients, it is not the substantial regression seen in other diabetic microvascular complications.[22] Potential risk factors of disease progression include hyperlipidemia, hypertension, and proteinuria.[21] Continued ophthalmic follow-up is necessary as some patients will still have progression of their disease. In particular, the most critical time for progression of retinopathy is the first year of metabolic normalization. What remains to be elucidated is the effect on diabetic retinopathy if patients could be operated on earlier in the course of their disease.[24]

Diabetic focal neuropathy and polyneuropathy are common complications of diabetes mellitus and often affect both the autonomic and somatic nervous systems. Polyneuropathy, a diffuse involvement of autonomic, sensory, and motor systems, is disabling and is the most common neurologic complication of diabetes. It is a contributing factor to foot ulceration. Chronic hyperglycemia generally is considered to be the most important factor in the development of diabetic neuropathy. In patients who underwent successful pancreas transplantation, neurologic evaluation demonstrated a general trend toward improvement in the motor and sensory nerve conduction studies at 1 year and in autonomic function at 5 years. In general, 60% showed improvement in motor testing, 50% showed improvement in sensory testing, and 30% to 45% showed improvement in autonomic nervous system testing.[2] Even with modest improvement, complete reversal of disease is not expected. Because of the continued risk of ulcers and amputations associated with neuropathy, pancreas transplant recipients should receive an annual foot examination by a trained health care provider. Smoking cessation is also recommended to reduce the risk of concomitant peripheral vascular disease.[21]

Quality of life is increasingly used to measure the success of an intervention such as transplantation. A consistent finding across quality-of-life studies in pancreas transplantation is improved perception about diabetes-specific issues such as satisfaction with diet flexibility and health management. The patient is relieved to be free from strict dietary restrictions, frequent blood sugar monitoring, and insulin therapy.[25]

The ultimate question is whether pancreas transplantation has any impact on overall survival. An analysis of data from the US Scientific Renal Transplant Registry indicates that SPK recipients can expect to live 15 years longer than type 1 diabetics on the wait list who do not receive a transplant. In addition, SPK recipients can expect to live 10 years longer than if they were to receive a KTA. In comparison with type 1 diabetics on the dialysis waiting list, it takes 101 days for SPK recipients to reach the same relative mortality risk because of the perioperative mortality.[2]

RECIPIENT CRITERIA

The most common indication for pancreas transplantation is C-peptide–deficient, insulin-dependent diabetes mellitus (type 1 diabetes mellitus.) A small percentage of transplants, roughly 5%, have been performed in recipients of type 2 diabetes. Similar patient survival rates and posttransplant insulin-independence rates have been reported.[1] The procedure has rarely been performed in cases of abnormal exocrine function. The considerations in patient selection entail criteria specific to the pancreas transplant and common to all potential recipients. In most centers, uncorrected cardiovascular disease is the major contraindication to pancreas transplantation. Other generally accepted contraindications include age greater than 60 years, active smoking, a left ventricular ejection fraction less than 30%, and significant obesity. Other common contraindications true to all organ recipients include untreated psychiatric illness, substance abuse, noncompliance, active infection, and recent malignancy.[18] Finally, recipient vascular disease is a risk factor for poorer patient survival. Iliac artery stenosis may complicate the technical procedure. For potential recipients with signs or symptoms of peripheral vascular disease, the adequacy of the peripheral vascular system should be assessed.[1]

Despite the current excellent results of SPK transplantation and the increasing knowledge of the many potential benefits of the procedure, African Americans have represented a disproportionately low percentage of all patients undergoing SPK transplantation. It has been suggested that previous reports of lower success rates in African American renal transplant recipients may be responsible for a perception that African Americans are at higher immunologic risk, possibly resulting in a reluctance to perform the more complex SPK procedure in these patients. It is also likely that the lower rate of SPK transplants in African Americans is related to ethnic disparities in access to the procedure. Although the data on SPK transplants in this population are limited because of the small numbers, there is some evidence of higher acute rejection. As a result, aggressive immunosuppression and a low threshold for biopsies should be utilized when African Americans undergo transplantation.[26]

DONOR CRITERIA

Because of the potential morbidity associated with pancreas transplantation, centers have been more conservative in the selection of pancreas donors than of donors of other intra-abdominal organs. This has undoubtedly contributed to the increasing waiting lists for pancreas transplantation. The traditional criteria restrict pancreas donors to ages of 10 to 45 years and weights of 30 to 80 kg. Hemodynamically unstable patients and those requiring large doses of vasopressors are generally excluded. Absolute contraindications to the use of a pancreas for transplantation have been donors with type 1 diabetes mellitus, a preinjury insulin requirement, morbid obesity, or significant pancreatic or duodenal trauma as well as the general organ contraindications of human immunodeficiency virus, active infection, and malignancy. Visual inspection and palpation of the organ by a transplant surgeon are essential. The overall quality and anatomy of the pancreas are assessed by the procurement team to identify possible fibrosis, fatty infiltration, evidence of severe trauma, saponification, and cysts, which are all unfavorable features. Pancreatic edema itself is not a contraindication.[18]

Unfortunately, the number of diabetic patients on the waiting list for pancreas transplantation continues to be greater than the number of available donor organs. Given this disparity, it has become clear that alternative solutions for donors must be identified. The shortage of cadaveric donors has prompted the use of cadaveric organs from pediatric donors. Recent studies have reported good long-term results using pediatric kidneys in adult recipients, but pediatric donors have rarely been used for pancreas transplantation because of concerns about islet cell mass, increased technical demands, and the possible risks of vascular thrombosis. An analysis of data from the University of Wisconsin group showed that the results of SPK transplants performed with pediatric donors whose weight, age, and body mass index were greater than a chosen threshold were excellent, with no increase in graft failure or thrombosis. The more liberal use of these donors as young as 4 could substantially increase the donor pool.[27]

Older age is a particularly important barrier to the acceptance of donors for pancreas transplants. The aging of the population and deceased donor pool raises the question of what age thresholds define acceptable rather than ideal deceased donors. One must balance the age of the donor organ with the increased mortality of waiting on the list for an ideal organ. Although old donors increase expected survival in comparison with dialysis, studies have shown that the acceptance of an old donor confers a higher risk of graft failure. Whether organs from old donors are more susceptible to ischemia reperfusion injury or trigger an additional release of cytokines and inflammatory and/or immunological mediators remains uncertain. The use of older donors should perhaps be restricted then to those patients who have decreased access to young donors on the basis of sensitization, rarer blood types, or listing in regions with long waiting list times.[28]

Organs donated after cardiac death (DCD) are being used to expand the kidney and liver pool and are now being explored as a source of organs for pancreas transplantation. DCD donors are hemodynamically stable individuals who, following a decision by their families to withdraw care and consent to organ donation, are extubated in the operating room. This is a planned event in which the procurement team is present to recover the organs rapidly, therefore limiting warm ischemia. There is a mandatory "no touch" period of 5 minutes after cardiac arrest in order to declare a patient clinically dead before the initiation of organ recovery. Recent data suggest that good patient and graft outcomes can support the use of selected DCD donors for pancreas transplantation. Recipients of SPK transplants from DCD donors had a shorter waiting time than recipients of SPK transplants from brain dead donors. The kidney allografts from DCD donors did have a higher incidence of delayed graft function (DGF). DCD donors should be controlled, young, lean, and hemodynamically stable. Organ procurement should be done by a well-trained pancreas team to minimize the warm ischemia time. The recipient operation should be coordinated to minimize the cold ischemia time.[29]

Efforts to expand the donor pool have even included live-donor pancreas transplantation. This involves the use of a segmental pancreas graft from a living donor. Live-donor pancreas transplantation is performed with the hopes of conveying a potential immunologic advantage and a decreased waiting time. This has been done in small numbers with no mortality but significant donor and recipient risks.[18]

PROCUREMENT OPERATION

After opening the lesser sac, the gastrocolic ligament is divided. The pancreas is closely examined and palpated. The right colon and duodenum are mobilized to the left. The right gastroepiploic and pyloric vessels are ligated and divided. The nasogastric tube is advanced into the second portion of the duodenum and instilled with a solution of saline, betadine, and amphotericin. The nasogastric tube is then withdrawn to the fundus of the stomach. The short gastric vessels are divided, and this allows the stomach to be separated from the spleen. The transverse colon is mobilized completely by ligation of the middle colic vasculature. The stomach then is divided proximally to the pylorus. The fourth portion of the duodenum is divided as well. While the spleen is carefully retracted anteriorly, the splenonephric and renophrenic attachments are divided. The liver is mobilized, and the aorta, vena cava, and inferior mesenteric vein are isolated. The gallbladder is emptied, and the bile duct is flushed. The supraceliac aorta is isolated, and heparin is given intravenously. Cannulas are placed into the aorta and inferior mesenteric vein, and preservation solution is instilled. The surface is cooled with ice slush. The thoracic organs then are procured, and after division of the gastroduodenal artery, portal vein, and splenic artery, the liver is removed. The pancreas with the spleen attached is removed after division of the superior mesenteric artery at the aorta. The allograft should be kept ice-cold and sterile in preservation solution until it is ready to be prepared on the backbench.[2]

The backbench preparation of a pancreatic allograft requires very careful preparation and surgical technique. The distal and proximal duodenum must be shortened to the proper length. This shortening is performed while the common bile duct is probed to ensure that the ampulla is in the center of the duodenal segment and is not compromised. The ends of the duodenum are then oversewn. The bile duct is ligated. A Y-graft is constructed with donor common iliac external iliac internal iliac artery bifurcation. The internal iliac artery is anastomosed to the splenic artery in an end-to-end fashion; then, the external iliac artery is anastomosed to the superior mesenteric artery. The inferior mesenteric vein is ligated. Then, the extraneous tissue at the periphery of the gland is ligated and removed. The hilum of the spleen is dissected carefully, and the splenic artery and vein are ligated. The stapled mesenteric vasculature is oversewn. The portal vein is then dissected free from the surrounding gland. Finally, the gland is tested for vascular integrity.[2]

SURGICAL PROCEDURE

The pancreas allograft is placed intra-abdominally and preferentially in the right lower quadrant. There are 2 major technical decisions to be made during pancreatic graft implantation: how to manage exocrine secretions and how to manage venous outflow. There are 3 options in managing the exocrine secretions of the pancreas: obstruction of the pancreatic duct, bladder drainage, and enteric drainage. Duct obstruction induced acute pancreatitis and was quickly abandoned.[30] The initial complications with enteric drainage of pancreatic exocrine secretions led to the development of bladder drainage. Before 1996, more than 90% of pancreas transplants were bladder-drained. The technique involves a side-to-side anastomosis between a segment of the transplant duodenum and the bladder. A major theoretical advantage of bladder drainage is that it facilitates the detection of pancreas rejection by measuring urinary amylase levels. However, in the presence of a simultaneous kidney transplant, an elevation of serum creatinine can be used as an early sign of rejection. Bladder drainage also avoids an enteric anastomosis and the associated contamination with enteric contents.[18] Exocrine drainage of the pancreas graft on the bladder is associated with the loss of large quantities of bicarbonate-rich pancreatic secretions in the urine, leading to metabolic acidosis, hyponatremia, and volume depletion.[1] The downside of bladder drainage is the development of complications requiring enteric conversion in approximately 25% of patients by 5 years after transplantation. The common complications and potential indications for conversion are urethritis, hematuria, recurrent urinary tract infections, metabolic acidosis, bladder stones, early anastomotic leaks, and reflux pancreatitis. For recipients of a bladder-drained pancreas, the rate of conversion to enteric drainage is 9% at 1 year and 17% at 3 years.[13] Enteric conversion is not a benign procedure, with as many as 15% of patients losing their grafts to rejection after conversion.[18]

Enteric drainage as a primary procedure is performed as a side-to-side duodenojejunostomy at 30 to 60 cm from the ligament of Treitz or as a Roux-en-Y anastomosis. The benefit of enteric drainage is the avoidance of urologic complications.[18] There are fewer urinary tract infections or episodes of dehydration and a lower rate of pancreatitis than with bladder drainage, and there are few, if any, significant metabolic derangements. The major specific problem with enteric drainage is the risk of more severe complications from the development of an enteric leak.[2] With improvements in surgical techniques and advances in clinical immunosuppression, the use of enteric drainage has increased again, but bladder drainage is still used by many centers.

Most pancreatic allografts have been transplanted like kidneys: heterotopically in the pelvis with the iliac vasculature for inflow and outflow. Advantages of this approach include lower rates of allograft thrombosis, easier access for percutaneous biopsy, and the ability to use either the bladder or intestine for drainage of exocrine secretions. This approach, however, does result in peripheral hyperinsulinemia with portal hypoinsulinemia because the first-pass effect of hepatic degradation is lost.[2] The actual clinical impact of this peripheral hyperinsulinemia remains uncertain. Down-regulation of insulin receptors in muscle and adipose tissue has been shown in SPK recipients. Peripheral hyperinsulinemia is known to alter lipoprotein metabolism and to have detrimental effects by being proatherosclerotic.[18]

Portal drainage is more physiologic. The first pass of insulin in the liver improves hepatic metabolism of glucose and prevents insulin resistance. Portal delivery of insulin can also prevent the potential metabolic complications of systemic hyperinsulinemia.[30] Portal venous drainage results in normal insulin levels, and lipoprotein metabolism is better than it is with systemic venous drainage. The rate of allograft thrombosis is somewhat higher, however, and access is more difficult if percutaneous biopsy is necessary. In addition, drainage must be enteric because of the cephalad placement of the donor duodenum.[2] However, to further confound matters, some studies have shown that antigen presentation to the liver via the portal vein produces systemic hyporesponsiveness, thus conveying an immunologic advantage and decreased incidence of rejection.[30] Despite the theoretical physiologic and immunologic superiority of portal-enteric drainage, the real clinical impact remains unproven, and its use is sporadic and center-specific.[18]

COMPLICATIONS

Results with pancreas transplantation have improved significantly during the last 25 years, in part because of improvements in antibiotics, surgical techniques, and immunosuppression. A number of early postoperative complications still can occur with pancreas transplantation. These complications include bleeding, infection, rejection, pancreatitis, posttransplant diabetes, and allograft thrombosis. The rate of technical complications after pancreas transplantation is higher than that after most other solid-organ transplantations. Technical failure rates (as opposed to immunologic failure rates) have been reported to be highest in SPK recipients, followed by PAK and PTA recipients.[2]

Graft thrombosis is well recognized and is the most common early cause of loss of a pancreatic allograft. It is a devastating complication that can occur in as many as 5% to 10% of pancreatic transplant recipients. Seventy percent of the time, graft thrombosis occurs in the first week after transplantation. Both recipient and donor factors can lead to graft thrombosis.[2] Type 1 diabetes mellitus has been shown to itself result in a hypercoagulable state involving varying components of the clotting cascade. In addition, hyperlipidemia, commonly associated with diabetes, further contributes to hypercoagulability. A thromboelastogram confirms this hypercoagulable state.[31] In addition, studies have demonstrated a high incidence of prothrombotic disorders.[32] Recipient factors are ones that decrease blood flow to the allograft and include pancreatitis, hypotension, acute rejection, and reperfusion injury. Donor factors include older age, longer cold ischemia times, and death resulting from cerebrovascular events. Thrombosis of a pancreatic allograft may occur in the arterial or venous system. Prevention and postoperative vigilance are the main measures for addressing the risk of allograft thrombosis. Anticoagulation is used by some centers to prevent graft thrombosis but increases the risk of postoperative bleeding. Liberal use of Doppler ultrasound imaging should be employed for any indication of early graft dysfunction.[2]

A small proportion of grafts, 3% to 5%, in the absence of technical problems never function sufficiently to induce insulin independence in the recipients (primary nonfunction) and are by definition early failures.[13] DGF occurs after many pancreas transplants but is poorly characterized. There is no well-accepted definition for DGF, but it can be considered the need for insulin on a regular basis after pancreas transplantation. DGF likely occurs because of both donor and recipient factors but usually implies some degree of graft injury. A better term may be glucose intolerance, as opposed to DGF, because the graft may be functioning adequately but the insulin requirement may be due to steroid and calcineurin inhibitor use. Some factors have been implicated, particularly donor age >45 years, perhaps as a result of decreased functional mass or preexisting atherosclerotic lesions of the pancreas leading to impaired microcirculation and early ischemia reperfusion injury. DGF does not seem to have a significant impact on long-term graft survival.[33]

Pancreatitis of an allograft, detected postoperatively by a transient rise in serum amylase or lipase levels, tends to occur to some degree in all recipients. These elevations usually are temporary and are without great clinical significance. Pancreatitis typically occurs as a result of cold ischemic storage and reperfusion injury or from handling during organ recovery and implantation and usually is without clinical consequence, is self-limited, and resolves with conservative therapy. Patients typically present with low-grade fever, elevated serum amylase levels, and graft tenderness. The treatment for transplant pancreatitis is generally supportive, but occasionally peripancreatic fluid collections can occur, and these collections may be high in exocrine secretions. Patients may require laparotomy, evacuation of collections, and pancreatic debridement if abdominal imaging suggests evidence of necrosis or subsequent infection.[2]

Duodenal segment leaks in the bladder-drained pancreas recipient most often occur in the first 3 months after transplantation and usually present with the acute onset of abdominal pain and elevation of serum amylase levels. Treatment is generally nonoperative with prolonged Foley catheter bladder drainage. Resistant cases may require exploration and closure of the leakage site or enteric conversion.

Enteric exocrine drainage is the most popular method of drainage today. Early duodenal segment leaks tend to result from technical complications or ischemia. Late duodenal leaks tend to be caused by rejection, infection, or ischemia of the duodenal staple line. Leaks do not result in an alteration of endocrine function, but patients present with elevated white blood cell counts, graft tenderness, and fever, and leaks generally lead to a pancreaticocutaneous fistula or peripancreatic abscess. These developments are particularly serious because of the spillage of succus entericus within the abdomen. Computed tomography imaging and percutaneous drainage usually demonstrate a mixed infection of bacteria and fungus. Broad-spectrum antibiotics are essential, and surgical exploration should be performed. At laparotomy, decisions must be made regarding the extent of the infection, its potential for clearance, and the potential need for removal of the pancreatic allograft. Treatment of the infection, if inadequate, will lead to organ failure, sepsis, and death of the recipient.[2]

Intra-abdominal infections are much more common after pancreas transplantation than after kidney transplants and represent a significant cause of mortality if inadequately treated. Conservative therapy with percutaneous drainage and intravenous antibiotics may be enough, although sometimes operative exploration and drainage may be necessary. A life-threatening complication of a mycotic aneurysm can occur in the setting of an untreated peripancreatic abscess.[2]

Early gastrointestinal bleeding, typically from the suture line of the enteric or bladder anastomosis, may occur. This bleeding can be caused by ischemia/reperfusion injury of the duodenal mucosa or by a bleeding vessel at the suture line of the anastomosis. When heparin is used for anticoagulation postoperatively, bleeding can be evident by either a fall in the hematocrit or melenic stool. Bleeding tends to stop with the cessation of anticoagulation. In bladder-drained patients, however, bladder irrigation may be necessary, and cystoscopy may be needed to remove larger clots.

The overall incidence of posttransplant diabetes mellitus is 17%. Factors associated with the development of persistent diabetes mellitus after pancreas transplantation include an increased pretransplant body mass index, a higher pretransplant daily insulin dose, and the occurrence of acute rejection. The development of persistent posttransplant diabetes mellitus is certainly multifactorial, and this entity can resemble both insulin-deficient and insulin-resistant states.[34] Normal-to-elevated C-peptide levels in these patients with hyperglycemia have been demonstrated. This may be secondary to exogenous weight gain or to the toxic effects of tacrolimus. Therefore, pancreas transplant recipients with hyperglycemia should have their C-peptide measured and undergo a pancreas biopsy to distinguish between rejection, recurrence of type 1 diabetes, and occurrence of type 2 diabetes. Therapy should then be directed at the appropriate pathologic condition.[21]

Immunosuppression carries with it an increased risk of infection: bacterial, fungal, and viral. Despite improvements in prophylactic therapy, intravenous ganciclovir and its oral equivalent valgancyclovir, cytomegalovirus (CMV) infection and disease remain significant causes of morbidity and mortality following pancreas transplantation. Pancreas transplant recipients appear to be at higher risk of developing CMV disease than other solid-organ transplant recipients. The causes are multifactorial but include an increase in the use of potent T-cell–depleting induction agents, a higher than expected number of CMV-seronegative recipients receiving organs from seropositive donors, and a high incidence of concomitant medical comorbidities. The overall incidence of CMV disease in patients undergoing SPK transplantation using standard immunosuppression is approximately 13% to 17%. Furthermore, CMV is associated with increased mortality following SPK transplantation, higher rates of rejection, and an increased risk of non-CMV infectious complications.[35]

Polyomaviruses are common in the adult population, with more than 80% of adults demonstrating serologic evidence of past exposure. In healthy individuals, a latent polyomavirus infection does not appear to result in any health consequences as the virus can establish latency in the uroepithelium. Polyomavirus-associated nephropathy, also known as BK virus nephropathy, is becoming recognized as an increasingly important cause of allograft dysfunction and loss in renal transplant recipients. The clinical manifestations of BK in renal allograft recipients include asymptomatic infection, ureteral strictures, transient impairment of function, and irreversible allograft failure. Once diagnosed, graft loss has been reported to be as high as 45% at 6 months. Much less is known about the clinical course of the disease in SPK transplant recipients. It is not thought that the pancreas allograft is vulnerable directly; however, this has not been clearly established, and the treatment of such infections in the renal allograft, typically with reduction of immunosuppression, places the pancreas allograft at risk for rejection.[36]

The risk of malignancy is also increased with immunosuppression post-transplantation. Careful surveillance is essential as in recipients of other solid-organ transplants. The risk of skin cancer may be as high as 50%, especially if there is increased sun exposure. Posttransplant lymphoproliferative disorder (PTLD) is a spectrum of diseases of the lymphoid cell lineage commonly associated with primary infection by Epstein-Barr virus in the context of posttransplant immunosuppression. The incidence in transplant recipients in general ranges from 0.8% to 20%. Analysis of PTLD in pancreas transplants is limited. The incidence appears to be comparable to that of recipients of other abdominal solid-organ allografts, but the clinical course seems to be markedly more aggressive, and survival is decreased. Pancreas recipients appear to have earlier onset PTLD as well as a higher proportion of T-cell lesions. A high index of suspicion is essential.[37]

REJECTION AND ROLE OF PANCREAS BIOPSIES

The increase in the pancreas allograft survival rates in recent years is the result of improved surgical techniques, adoption of more potent immunosuppression, careful management of patients, and aggressive monitoring of graft function after transplantation. Pancreas allografts are monitored by laboratory tests that assess the adequacy of the endocrine and exocrine functions and by radiologic techniques that determine blood flow into the graft. The specificity of these markers in establishing the diagnosis of acute rejection is suboptimal. Clinical features for the diagnosis of acute rejection, such as elevated serum amylase or lipase levels, a 50% decrease in urinary amylase for bladder-drained grafts, unexplained fever, and hyperglycemia have been associated with a positive predictive value of only 75%. Hyperglycemia can result from multiple etiologies and is frequently a sign of late or severe acute rejection. Progressive insulin dependency is often a sign of chronic rejection. Pancreas allograft biopsy is the gold standard for the diagnosis of acute rejection. Kidney biopsy, for easier accessibility, has been used as a surrogate for a pancreas biopsy in patients with an SPK transplant.[38]

The standard procedure to obtain a pancreas allograft biopsy is the percutaneous needle biopsy technique. Biopsies are performed under computerized tomography scanning or ultrasound. Biopsies can also be obtained laparoscopically. Biopsies are generally safe, with a reported incidence of bleeding or inadvertent biopsy of other organs of 2.8%. Surgical intervention is required in only 1.2% of patients who undergo a biopsy.[38]

Similar to that in other allografts, acute rejection in the pancreas is characterized by inflammatory cellular infiltrates associated with features of target tissue injury. Venous endothelitis is important to diagnose acute rejection and a requirement for the diagnosis of very mild rejection when inflammation is restricted to the interlobular septae. Higher grades of rejection are determined by the involvement of the arteries and by the diffuse and destructive nature of the inflammatory reaction. The histologic grades correlate with rejection severity as determined by the clinical degree of graft dysfunction and treatment response. Diffuse moderate-to-severe acinar inflammation with necrosis, intimal arteritis, and venulitis correlates with graft failure.[38]

Antibody-mediated acute rejection of pancreas allografts is poorly understood, mostly because of the inability to make the diagnosis. Deposition of complement fragment C4d in allografts, in conjunction with evidence of tissue and microvascular injury, and/or detection of anti–human leukocyte antigen or anti–donor-specific antibodies is diagnostic for antibody-mediated rejection in renal transplants and is now being looked at in pancreas transplants. C4d is a complement split product generated during antibody-mediated immune response. It covalently binds to the capillary endothelium at the time of injury. On the other hand, the histologic features of chronic rejection are the result of vascular sclerosis and progressive fibrosis, with a loss of functioning structures. Vascular lesions of chronic rejection include fibrocellular intimal hyperplasia with or without foam cells and lumen compromise.[38]

IMMUNOSUPPRESSION

When azathioprine and steroids were the cornerstone of maintenance therapy in pancreas transplantation, the acute rejection rates ranged from 50% to 80%. Almost all possible combinations of the immunosuppressants clinically available have been used in pancreas transplants. To date, most centers favor a quadruple-drug regimen: an induction agent, tacrolimus, mycophenolate mofetil, and steroids.

In the early 1990s, cyclosporine accounted for virtually 100% of the calcineurin inhibitor use. Since the mid 1990s, the most popular oral maintenance immunosuppressant combination has been tacrolimus (FK506 or Prograf) and mycophenolate mofetil (CellCept). Tacrolimus is more effective than cyclosporine in preventing moderate or severe kidney or pancreas rejection after pancreas transplantation. However, a higher incidence of posttransplant diabetes mellitus can be observed in solid-organ transplant patients with tacrolimus-based immunosuppression.[30]

Antimetabolites continue to be used in the vast majority of regimens for pancreas transplantation. Mycophenolate mofetil has replaced the use of azathioprine. Mycophenolate mofetil has been shown to be effective in pancreas transplantation by increasing short-term graft survival and lowering rejection rates.[18] Sirolimus is emerging as a novel immunosuppressant for pancreas transplantation. Sirolimus offers the advantage of not being diabetogenic.

Because of the prevalence of early rejection and the difficulty of its diagnosis, induction therapy is almost routinely included in immunosuppressive protocols for pancreas transplant recipients. Induction therapy is a complementary form of immunosuppression involving antilymphocyte antibody pharmacologics that are parenterally administered for a short course immediately peri-transplant and posttransplant. The rationale for using induction therapy pertains to the agents' potent anti–T-cell immunosuppressive properties. In this context, induction therapy is used in conjunction with maintenance agents for the purpose of minimizing the risks of early rejection episodes.[39] The agents available can be divided into 2 groups: (1) T-cell–depleting antibodies, polyclonal (eg, thymoglobulin) or monoclonal (eg, OKT3 and Campath), and (2) nondepleting antibodies (monoclonal anti-CD25 directed antibodies, eg, Zenapax and Simulect). The advantages of these agents in preventing acute rejection need to be balanced against the potential for infectious complications and the high cost of these agents.

Corticosteroid avoidance protocols are gaining acceptance in the field of renal transplantation because of favorable long-term safety and efficacy outcomes. Steroids have many metabolic side effects, including peripheral insulin resistance, which reduces the benefit of pancreas transplantation and possibly leads to the reintroduction of exogenous insulin treatment. In addition, steroids may lead to osteoporosis, osteonecrosis, cataracts, hyperlipidemia, weight gain, and hypertension. It is important to weigh these potential risks against the risk of organ rejection. The most influential factor promoting patient survival in SPK recipients is a functioning renal allograft, and the loss of a kidney graft has a more negative impact on the patient survival rate than the loss of a pancreas graft. Furthermore, SPK recipients that have to return to dialysis are at an increased risk of infection because of the obvious need to continue immunosuppression therapy to maintain their pancreas allograft.[40] There are several centers that are exploring the best regimens to allow for steroid-free maintenance therapy in pancreas transplantation. Minimization immunosuppressive strategies may offer another avenue to reduce the incidence of CMV in SPK patients. The elimination of the use of corticosteroids appears to provide another approach to the reduction of CMV disease.[35]

ISLET TRANSPLANTATION

Despite substantial improvements in insulin therapy and the care of patients with type 1 diabetes mellitus, a subgroup of patients is still disabled by refractory hypoglycemia. Cell-based therapy with islet transplantation offers a possibility for improved glycemic control without the morbidity and mortality of a surgical procedure. The past 3 decades have witnessed substantial progress in islet transplantation. In 2000, Shapiro *et al.*[41] reported their initial findings with up to a year of follow-up in 7 consecutive subjects treated with an infusion of an adequate mass of freshly prepared islets from 2 or more pancreases from deceased donors combined with glucocorticoid-free immunosuppressive therapy. In all 7 subjects, insulin independence was achieved with tight glycemic control and correction of glycated hemoglobin levels. This treatment became known as the Edmonton protocol.[41] Sustained C-peptide production and successful insulin independence were achieved and then replicated in other transplant centers. An international, multicenter trial confirmed the previous experiences with the Edmonton protocol, demonstrating the reproducibility and benefits of islet alone transplantation in type I diabetes mellitus with unstable glycemic control. The trial succeeded in standardizing pancreas selection, islet processing, recipient selection, and posttransplant care.[41]

Islets are prepared after human cadaveric pancreases are removed from multiorgan donors following in situ vascular flushing and transported to clinical islet isolation laboratories. The pancreatic duct is cannulated, and the pancreas is enzymatically and mechanically dissociated. The recipient is brought to the department of radiology, and interventional radiologists cannulate the portal vein. The islets are allowed to infuse under gravity pressure. Immunosuppression has customarily been induction therapy plus tacrolimus and sirolimus. The major acute complications of the percutaneous transhepatic approach are serious bleeding and portal vein thrombosis. The chronic complications of the immunosuppressive therapy must also be weighed against freedom from hypoglycemia and glycemic lability.[42]

The National Institute of Diabetes and Digestive and Kidney Diseases founded the Collaborative Islet Transplant Registry to expedite progress and promote safety in islet cell transplantation through the voluntary collection, analysis, and communication on all islet transplants performed in North America. Nineteen islet programs have contributed through 2005, providing extensive data on 118 islet transplant alone (ITA) recipients (receiving 231 islet infusions), 19 islet after kidney recipients (receiving 34 islet infusions), and no simultaneous islet kidney recipients.[43]

Failure rates for islet allotransplantation in nonuremic type 1 diabetics are higher than those for whole-organ pancreas transplantation at the current time. Reported outcomes at 6 months following the patients' last islet infusion showed that 67% were insulin-independent, but this rate dropped to 58% at 1 year. For recipients who were still on insulin at 6 months after their last infusion, there was a 56% mean reduction in exogenous insulin injected in comparison with the baseline. In addition, there was a striking decrease in the occurrence of severe hypoglycemic events after even just 1 islet infusion, from over 82% in the 6 months before islet infusion in the ITA patients to only 2.5% in the 30 days after infusion and 0% in the subsequent 5 months. Although sustained insulin independence is not reliably achieved with ITA, there is an important quality-of-life improvement with even partial function of islets due to the absence of hypoglycemia.[43]

The mechanisms of islet failure are likely multifactorial. There may be inadequate islet cell mass or impaired function. It is likely that many islets are lost during purification and engraftment.[42] One method of early islet graft failure has been hypothesized to be the induction of liver ischemia by islet thrombi with the resultant expression of proinflammatory cytokines. The resultant apoptosis and necrosis of surrounding liver tissue could result in promoting subsequent islet deterioration. The importance of antibodies to nonself human leukocyte antigens in islet transplantation outcomes has also been recently elucidated. It has been shown that preexisting anti–human leukocyte antigen antibodies (sensitization) may be one of the factors that could be detrimental to the function of transplanted islets.[43]

CONCLUSION

Diabetes mellitus is a chronic disease often leading to debilitating microvascular and macrovascular complications. Normalization of glucose homeostasis is essential for reducing the cardiovascular morbidity and mortality. There has been a great deal of excitement and investigation in islet transplantation, but at the current time only whole-organ pancreas transplantation can be counted on for long-term exogenous insulin independence and normalization of hemoglobin A1c. SPK transplantation is the preferred treatment option for diabetics with approaching or coexisting end-stage renal disease. The option of a cadaver PAK transplant may be considered after living donor kidney transplantation. Finally, PTA may be offered to diabetics with good renal function but difficulty with glycemic control. The benefits for all are stable glucose control and improved quality of life. The risks are surgical morbidity and the long-term risks of immunosuppression. A final issue is the imbalance between supply and demand and the overwhelming shortage of organs.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. van der Boog JM, Ringers J, Paul LC, et al. Simultaneous kidney-pancreas transplantation: the preferred option for patients with type I diabetes mellitus and approaching end-stage renal disease. *Transplant Rev* 2004; 18: 129.
2. Lipshutz GS, Wilkinson AH. Pancreas-kidney and pancreas transplantation for the treatment of diabetes mellitus. *Endocrinol Metab Clin North Am* 2007; 36: 1015–1038.
3. Rayhill SC, D'Alessandro AM, Odorico JS, et al. Simultaneous pancreas-kidney transplantation and living related donor renal transplantation in patients with diabetes: is there a difference in survival? *Ann Surg* 2000; 231: 417–423.
4. Stadler M, Auinger M, Anderwald C, et al. Long-term mortality and incidence of renal dialysis and transplantation in type 1 diabetes mellitus. *J Clin Endocrinol Metab* 2006; 91: 3814–3820.
5. 2004 Annual Report of the U.S. Organ Procurement and Transplantation Network and the Scientific Registry of the Transplant Recipients: Transplant Data 1994–2003. Available at: http://www.optn.org. Accessed June 2008.
6. Andreoni KA, Brayman KL, Guidinger MK, et al. Kidney and pancreas transplantation in the United States, 1996–2005. *Am J Transplant* 2007; 7(pt 2): 1359–1375.
7. Gruessner RW, Sutherland DE, Gruessner AC. Mortality assessment for pancreas transplants. *Am J Transplant* 2004; 4: 2018–2026.
8. Humar A, Ramcharan T, Kandaswamy R, et al. Pancreas after kidney transplants. *Am J Surg* 2001; 182: 155–161.
9. Dinavahi R, Akalin E. Preemptive kidney transplantation in patients with diabetes mellitus. *Endocrinol Metab Clin North Am* 2007; 36: 1039–1049.
10. Israni AK, Feldman HI, Propert KJ, et al. Impact of simultaneous kidney-pancreas transplant and timing of transplant on kidney allograft survival. *Am J Transplant* 2005; 5: 374–382.
11. Farney AC, Cho E, Schweitzer EJ, et al. Simultaneous cadaver pancreas living-donor kidney transplantation: a new approach for the type 1 diabetic uremic patient. *Ann Surg* 2000; 232: 696–703.
12. Larson TS, Bohorquez H, Rea DJ, et al. Pancreas-after-kidney transplantation: an increasingly attractive alternative to simultaneous pancreas-kidney transplantation. *Transplantation* 2004; 77: 838–843.
13. Gruessner AC, Sutherland DE. Pancreas transplant outcomes for United States (US) and non-US cases as reported to the United Network for Organ Sharing (UNOS) and the International Pancreas Transplant Registry (IPTR) as of June 2004. *Clin Transplant* 2005; 19: 433–455.
14. Luan FL, Miles CD, Cibrik DM, Ojo AO. Impact of simultaneous pancreas and kidney transplantation on cardiovascular risk factors in patients with type 1 diabetes mellitus. *Transplantation* 2007; 84: 541–544.

15. Jukema JW, Smets YF, van der Pijl JW, et al. Impact of simultaneous pancreas and kidney transplantation on progression of coronary atherosclerosis in patients with end-stage renal failure due to type 1 diabetes. *Diabetes Care* 2002; 25: 906–911.
16. Dieterle CD, Schmauss S, Arbogast H, et al. Pulmonary function in patients with type 1 diabetes before and after simultaneous pancreas and kidney transplantation. *Transplantation* 2007; 83: 566–569.
17. Coppelli A, Giannarelli R, Vistoli F, et al. The beneficial effects of pancreas transplant alone on diabetic nephropathy. *Diabetes Care* 2005; 28: 1366–1370.
18. Di Carlo A, Odorico JS, Sollinger HW. Pancreas transplantation: an overview. *Transplant Rev* 2004; 18: 1.
19. Coppelli A, Giannarelli R, Boggi U, et al. Disappearance of nephrotic syndrome in type 1 diabetic patients following pancreas transplant alone. *Transplantation* 2006; 81: 1067–1068.
20. Giannarelli R, Coppelli A, Sartini M, et al. Effects of pancreas-kidney transplantation on diabetic retinopathy. *Transpl Int* 2005; 18: 619–622.
21. Mai ML, Ahsan N, Gonwa T. The long-term management of pancreas transplantation. *Transplantation* 2006; 82: 991–1003.
22. Chow VC, Pai RP, Chapman JR, et al. Diabetic retinopathy after combined kidney-pancreas transplantation. *Clin Transplant* 1999; 13: 356–362.
23. Giannarelli R, Coppelli A, Sartini MS, et al. Pancreas transplant alone has beneficial effects on retinopathy in type 1 diabetic patients. *Diabetologia* 2006; 49: 2977–2982.
24. Pearce IA, Ilango B, Sells RA, Wong D. Stabilisation of diabetic retinopathy following simultaneous pancreas and kidney transplant. *Br J Ophthalmol* 2000; 84: 736–740.
25. Sureshkumar KK, Patel BM, Markatos A, et al. Quality of life after organ transplantation in type 1 diabetics with end-stage renal disease. *Clin Transplant* 2006; 20: 19–25.
26. Rogers J, Baliga PK, Chavin KD, et al. Effect of ethnicity on outcome of simultaneous pancreas and kidney transplantation. *Am J Transplant* 2003; 3: 1278–1288.
27. Fernandez LA, Turgeon NA, Odorico JS, et al. Superior long-term results of simultaneous pancreas-kidney transplantation from pediatric donors. *Am J Transplant* 2004; 4: 2093–2101.
28. Salvalaggio PR, Schnitzler MA, Abbott KC, et al. Patient and graft survival implications of simultaneous pancreas kidney transplantation from old donors. *Am J Transplant* 2007; 7: 1561–1571.
29. Salvalaggio PR, Davies DB, Fernandez LA, Kaufman DB. Outcomes of pancreas transplantation in the United States using cardiac-death donors. *Am J Transplant* 2006; 6(pt 1): 1059–1065.
30. Ming CS, Chen ZH. Progress in pancreas transplantation and combined pancreas-kidney transplantation. *Hepatobiliary Pancreat Dis Int* 2007; 6: 17–23.
31. Burke GW III, Ciancio G, Figueiro J, et al. Hypercoagulable state associated with kidney-pancreas transplantation. Thromboelastogram-directed anti-coagulation and implications for future therapy. *Clin Transplant* 2004; 18: 423–428.
32. Wullstein C, Woeste G, Zapletal C, et al. Prothrombotic disorders in uremic type-1 diabetics undergoing simultaneous pancreas and kidney transplantation. *Transplantation* 2003; 76: 1691–1695.
33. Tan M, Kandaswamy R, Sutherland DE, et al. Risk factors and impact of delayed graft function after pancreas transplants. *Am J Transplant* 2004; 4: 758–762.
34. Dean PG, Kudva YC, Larson TS, et al. Posttransplant diabetes mellitus after pancreas transplantation. *Am J Transplant* 2008; 8: 175–182.
35. Axelrod D, Leventhal JR, Gallon LG, et al. Reduction of CMV disease with steroid-free immunosuppression in simultaneous pancreas-kidney transplant recipients. *Am J Transplant* 2005; 5: 1423–1429.
36. Lipshutz GS, Mahanty H, Feng S, et al. BKV in simultaneous pancreas-kidney transplant recipients: a leading cause of renal graft loss in first 2 years post-transplant. *Am J Transplant* 2005; 5: 366–373.
37. Paraskevas S, Coad JE, Gruessner A, et al. Posttransplant lymphoproliferative disorder in pancreas transplantation: a single-center experience. *Transplantation* 2005; 80: 613–622.
38. Gaber LW. Pancreas allograft biopsies in the management of pancreas transplant recipients: histopathologic review and clinical correlations. *Arch Pathol Lab Med* 2007; 131: 1192–1199.
39. Kaufman DB, Iii GW, Bruce DS, et al. Prospective, randomized, multi-center trial of antibody induction therapy in simultaneous pancreas-kidney transplantation. *Am J Transplant* 2003; 3: 855–864.
40. Gallon LG, Winoto J, Chhabra D, et al. Long-term renal transplant function in recipient of simultaneous kidney and pancreas transplant maintained with two prednisone-free maintenance immunosuppressive combinations: tacrolimus/mycophenolate mofetil versus tacrolimus/sirolimus. *Transplantation* 2007; 83: 1324–1329.
41. Shapiro AM, Ricordi C, Hering BJ, et al. International trial of the Edmonton protocol for islet transplantation. *N Engl J Med* 2006; 355: 1318–1330.
42. Ryan EA, Paty BW, Senior PA, et al. Five-year follow-up after clinical islet transplantation. *Diabetes* 2005; 54: 2060.
43. Pavlakis M, Khwaja K. Pancreas and islet cell transplantation in diabetes. *Curr Opin Endocrinol Diabetes Obes* 2007; 14: 146–150.

Mount Sinai Journal of Medicine 75:385–397, 2008

Type 1 Diabetes Mellitus: Primary, Secondary, and Tertiary Prevention

Jennifer Bollyky, MD,[1] Srinath Sanda, MD,[1,2] and Carla J. Greenbaum, MD[1]

[1] Diabetes Program, Benaroya Research Institute, Seattle, WA
[2] Department of Pediatrics, Seattle Children's Hospital, Seattle, WA

ABSTRACT

We have entered the era of clinical trials to prevent type 1 diabetes mellitus (T1DM). Before 1922, when insulin was first given to a patient with diabetes, a diagnosis of T1DM was considered a death sentence. Advances in treatment for subjects with diabetes are not yet sufficient to prevent the deleterious impact of diabetes on both day-to-day activities and the early morbidity and mortality still associated with the disease. We now understand a great deal about blood glucose regulation and potential health complications associated with long-term T1DM, but the mystery of why, or the pathogenesis of this devastating disease, remains elusive. Great strides toward unraveling this mystery have been made over the past several decades. Even without definitive answers, we are moving from the period of discovery and animal research to the era of clinical trials. In this review, we wish to convey the palpable excitement in the field. It is time to determine if we can safely change the course of T1DM. ***Mt Sinai J Med 75:385–397, 2008.*** © *2008 Mount Sinai School of Medicine*

Key Words: Type I diabetes, clinical trial, autoimmunity, prevention.

Although diabetes was known in ancient times, it was not until 1889 that Von Mering and Minkowski at the University of Strasbourg were able to demonstrate that the removal of the pancreas in dogs caused diabetes.[1] It then took 33 years before this discovery translated into therapy when Banting and Best identified, isolated, purified, and treated the first subject with insulin. It was not until the 1970s that evidence convincingly demonstrated that insulin-dependent diabetes was an immune-mediated disease. The evidence included measurements of autoantibodies to islet cells[2,3] (and subsequently insulin[4] and others[5–7]), pathology reports describing insulitis in pancreatic specimens,[8,9] and associations of the disease with genes that encode cell-surface antigen-presenting proteins [the human leukocyte antigen (HLA) system or HLA type].[10–12] In the 1980s, reports of islet cell and insulin antibodies together with deficiency of insulin secretion were noted in patients prior to the onset of clinical disease.[13] The Eisenbarth model of disease, describing a preclinical period of autoimmune beta cell destruction followed by overt disease, was widely accepted.[14,15] During the same era, advances in immunotherapy, reports and then standardization of measurements of additional autoantibodies, and the ability to change the course of disease in the 2 animal models (nonobese diabetic mouse and biobreeding rat) led to the development of clinical trials to do the same. As illustrated in Figure 1, increased understanding of the natural history of the T1DM disease process expanded the concept of prevention to include primary prevention, in which the goal is to stop the initiation of autoimmunity; secondary prevention, in which the goal is to stop the progression of immune-mediated beta cell destruction and prevent clinical disease; and tertiary prevention, with the aim of easing day-to-day management of diabetes and diminishing complications.

NATURAL HISTORY OF DISEASE

The incidence of T1DM in the general Caucasian population is about 0.3%. In contrast, the overall risk among family members is about 15 times greater. Thus, although the majority of individuals who develop T1DM have no family members with

Address Correspondence to:

Carla J. Greenbaum, MD
Diabetes Program
Benaroya Research Institute
Seattle, WA
Email: cjgreen@benaroyaresearch.org

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20054

© 2008 Mount Sinai School of Medicine

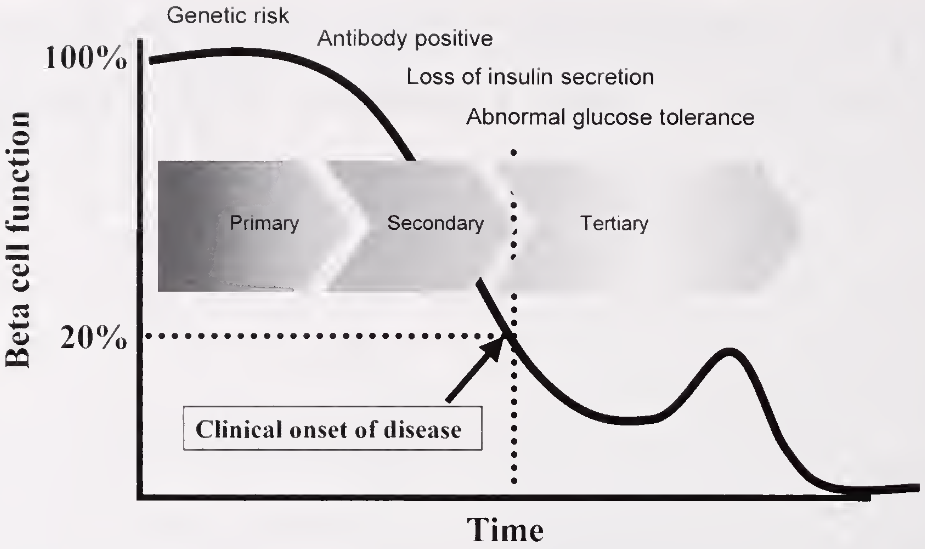


Fig 1. Stages of Type 1 Diabetes Prevention. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.].

disease, the majority of the data about the natural history of disease come from studies of family members. A large amount of such data has confirmed that a small percentage of individuals at genetic risk develop autoimmunity. This autoimmunity is detected first by the presence of autoantibodies and second by the loss of beta cell function assessed by impaired insulin secretion and subsequently abnormal glucose tolerance. Overt clinical diabetes occurs only after the remaining beta cell function is insufficient to prevent hyperglycemia. At clinical onset, beta cell function is further transiently depressed. For many individuals, a temporary improvement in glucose tolerance is seen within the first few months of diagnosis–the honeymoon period. The honeymoon period is characterized clinically as requiring decreased exogenous insulin and metabolically by improvements in both insulin action and secretion. Whether the honeymoon period is due to the resolution of the acute effects of metabolic decompensation on beta cells and resolution of insulin-sensitive tissues or the immune-mediated cytokine impairment of beta cell function is unknown. Post-diagnosis, immune-mediated beta cell dysfunction and destruction continue, as measured by decreased responses to beta cell insulin secretagogues.

This overall description of the T1DM disease process masks the heterogeneity evident in clinical practice. Some individuals progress to clinical T1DM in infancy; others do so during late adulthood. Whether this is because the disease process is more rapid in younger children or they start with less islet reserve is unknown. Post–clinical diagnosis, about 15% of individuals still have measurable beta cell function even 5 years later,[16–18] and recent data suggest that small amounts of insulin secretion may persist for decades in selected patients. Overall, however, although the vast majority of patients present with significant amounts of beta cell function at diagnosis, this is gradually and inevitably lost over time.

Although our current knowledge about the natural history of the disease allows for the selection of subjects for clinical trials, further understanding of the causes of the disease will need to take into account emerging information. For example, the incidence of T1DM is increasing worldwide. The increase is particularly evident in the very young.[19–23] Such an increase suggests environmental triggers during pregnancy or early life. Although such triggers have been previously explored, an ongoing study (The Environmental Determinants of Diabetes in the Young) joins the efforts of international investigators to better understand environmental causes of autoimmunity in newborns.[24] Another relatively recent observation is that there has been an apparent change in the HLA type among individuals being diagnosed with T1DM and that the increase in disease is often in those subjects with moderate-risk or low-risk genotypes.[25,26] Our developing understanding of the natural history of T1DM must also take into account the influence of lifestyle behaviors and obesity on the disease course. Along with standard assessments of the HLA type, autoantibody status, and metabolic tests, lifestyle data are being collected prospectively as part of another large international effort, the Type 1 Diabetes TrialNet Natural History Study. Both of these large international studies are also collecting DNA, RNA, and peripheral mononuclear blood cells for use by the research community in exploring the etiopathology of the disease.

PREVENTION STUDIES: RATIONALE AND FEASIBILITY

Studies of the primary, secondary, and tertiary prevention of T1DM are summarized in Table 1.

Primary Prevention

A completely safe, inexpensive, and easily administered therapy to all newborns is theoretically possible. The safety and feasibility bar would need to be quite high because one would be treating 300 infants to prevent 1 case of diabetes. Such a universal approach could be less expensive than a more targeted strategy as it would obviate the need for screening tests. The rationale for such primary prevention is obvious: the elimination of T1DM. However, because there is, as yet, no therapy that has been shown to prevent disease, enrollment in a clinical primary prevention trial currently requires more selective targeting. HLA testing of either the general population (ie, no family members with T1DM) or infants with relatives with the disease can identify those with high-risk or moderate-risk genotypes.[27–29] Among first degree relatives, about 20% of infants with high risk will develop autoimmunity within the first 2 years of life.[29] Although details of study design are beyond the scope of this review, if one expects the therapy being tested to reduce autoimmunity by a moderately large amount, current knowledge and technology demonstrate that primary prevention studies are feasible. This is clearly illustrated by the ongoing Trial To Reduce Insulin-Dependent Diabetes in the Genetically at Risk (TRIGR), which is testing whether early exposure to cow's milk influences diabetes risk. Epidemiology data both support[30] and refute[31] an effect of early introduction of cow's milk as a risk factor for islet autoantibody development. TRIGR is a multinational effort that has screened newborns for genetic risk and enrolled more than 2000 in a placebo-controlled, randomized trial.[32] TRIGR screening and enrollment were completed 1 year early, and the results of this study should be available within the next few years.

Another potential dietary intervention relies on data indicating that the early introduction of cereal[33] and gluten[34] is associated with islet autoimmunity. These observations led to the BabyDiet study, a pilot, primary prevention trial testing whether the development of autoantibodies differs in babies randomized to early (at 6 months of age) or late (12 months of age) introduction of gluten.[35] This study is underway in Germany.

Another pilot study, which began in 2007 and is called the Nutritional Intervention To Prevent Diabetes Study, has been designed to test the hypothesis that omega-3 fatty acids may protect against autoimmune disease. Epidemiological evidence supports this hypothesis, in that a change in Western diets from anti-inflammatory acids to proinflammatory acids over the past century[36] has coincided with an increase in the incidence of T1DM. Also, increased levels of proinflammatory prostaglandin synthase are seen in children with diabetes.[37] Furthermore, a population-based, case-controlled study has suggested that cod liver oil may be protective against the onset of T1DM.[38] In the TrialNet-sponsored Nutritional Intervention to Prevention Diabetes (NIP) primary prevention study, mothers with genetically at-risk fetuses are being given omega-3 fatty acid supplementation in the third trimester of pregnancy. These infants or others found to be at high genetic risk also receive supplements for the first several years of life.

There has been increasing attention paid to the intersection of epidemiologic data suggesting that vitamin D insufficiency is a risk factor for T1DM and basic science studies demonstrating multiple effects of vitamin D on immune and metabolic function.[39,40] Recent pilot studies have indicated that vitamin D can be safety administered even to the very young,[41] and a primary prevention trial may be considered by TrialNet in the near future.

These efforts all point to worldwide acceptance of primary prevention as the safest way to stop the disease and, moreover, demonstrate that such studies are feasible.

Secondary Prevention

The feasibility of secondary prevention studies was demonstrated during the past decade with the successful enrollment of 3 large, multicenter trials of family members identified as being at risk for progression from autoimmunity to clinical disease. Two of these studies were conducted in North America under the auspices of the Diabetes Prevention Trial–Type 1 Diabetes (DPT-1). In 1 study, islet cell antibody (ICA)–positive relatives with markedly abnormal insulin secretion or abnormal glucose tolerance were enrolled in a randomized

Table 1. *Type 1 Diabetes Mellitus Prevention Studies.*

Primary Prevention (Infants and Pregnant Women)

		Enrollment Status	Contact
NIP	Docosahexanoic acid capsules versus placebo	Open	DiabetesTrialNet.org 1-800-HALT-DM1
TRIGR	Bovine protein (cow's milk) or hydrolyzed casein formula	Closed	trigr.epi.usf.edu
BabyDiet	Early (6 months) or late (12 months) introduction of gluten	Closed	
Pre-POINT	Oral and nasal insulin	Not yet opened	JDRF.org

Secondary Prevention (Relatives of Individuals with T1DM)

		Enrollment Status	Contact
Oral insulin	Oral insulin capsules or placebo	Open	DiabetesTrialNet.org 1-800-HALT-DM1
INIT-II (Australia)	Nasal insulin	Open	DVDC dvdc.org.au
Natural History Study	Determine risk for entry into trials	Open	DiabetesTrialNet.org 1-800-HALT-DM1
Anti-CD3 in very high risk subjects	hOKT3γ1 (ala-ala)	Not yet opened	DiabetesTrialNet.org 1-800-HALT-DM1

Tertiary Prevention (Individuals with T1DM)

		Enrollment Status	Contact
Abatacept	CTLA4-Ig versus placebo	Open: ages 6–45	DiabetesTrialNet.org 1-800-HALT-DM1
ABATE study	hOKT3γ1 (ala-ala)	Open: ages 8–30	abatetrial.org
Thymoglobulin	ATG versus placebo	Open	immunetolerance.org
IL-2 and rapamycin	Phase I trial	Open: ages 18–45	immunetolerance.org
Anti-CD20	Rituximab versus placebo	Closed	DiabetesTrialNet.org 1-800-HALT-DM1
MMF/DZB	Three-arm trial: MMF/DZB, MMF/placebo, and placebo/placebo	Closed	DiabetesTrialNet.org 1-800-HALT-DM1
Anakinra	IL-1B receptor agonist versus placebo	Not yet opened	JDRF.org
Metabolic intervention	Closed-loop glucose/insulin and real-time glucose monitoring versus usual care	Not yet opened	DiabetesTrialNet.org 1-800-HALT-DM1
GAD65 in alum	rhGAD65 (alum) versus placebo	Not yet opened	DiabetesTrialNet.org 1-800-HALT-DM1

On-going or planned government or non-profit sponsored primary, secondary, or tertiary prevention trials. **Abbreviations:** ABATE, Autoimmunity-Blocking Antibody Therapy for Tolerance in Type 1 Diabetes; ATG, anti-thymocyte globulin; DVDC, Diabetes Vaccine Development Center; DZB, daclizumab; GAD65, glutamic acid decarboxylase 65; IL, interleukin; INIT-II, Intranasal Insulin Trial II; MMF, mycophenolate mofetil; NIP, Nutritional Intervention To Prevent Diabetes; Pre-POINT, Primary Oral/Intranasal Insulin Trial; rhGAD65, recombinant glutamic acid decarboxylase 65; T1DM, type 1 diabetes mellitus; TRIGR, Trial To Reduce Insulin-Dependent Diabetes in the Genetically at Risk.

trial. These high-risk individuals were expected to have a greater than 50% risk of progression to clinical disease within 5 years.[42] The other DPT-1 trial enrolled subjects with ICAs and insulin autoantibodies (IAAs) in a double-masked, placebo-controlled trial. These subjects were estimated to have a 25% to 50% 5-year risk of diabetes.[43] The European Nicotinamide Diabetes Intervention Trial (ENDIT) was another double-masked, placebo-controlled trial in antibody-positive relatives.[44] Although the intervention tested in each of these studies failed to delay or prevent the onset of disease, the trials demonstrated the feasibility of secondary prevention studies. Although large numbers of relatives needed to be tested (>100,000 in the DPT-1 trial), sufficient subjects were enrolled, and study compliance was extremely high. Moreover, these large studies confirmed the prediction algorithms. The 5-year risk of diabetes was found to be 60% in the high-risk DPT-1 group and 35% in the intermediate-risk group. Importantly, these studies also confirmed the benefit to subjects participating in such trials, in which they are closely monitored for the development of diabetes. Almost two-thirds of those who developed diabetes were diagnosed during routine follow-up visits when they were asymptomatic, and this prevented the morbidity often associated with diabetes onset.

Because of the large numbers of subjects involved, data from ENDIT and DPT-1 have proved to be valuable for additional insights into the natural history of disease progression. For example, subjects with abnormal glucose tolerance at the time of enrollment [eg, impaired fasting glucose (>110 mg/dL), indeterminate glucose tolerance (30-, 60-, or 90-minute glucose >200 mg/dL), or impaired glucose tolerance (2-hour glucose, 141–199 mg/dL)] were noted to have an extremely high risk of progression (~80% 5-year risk). Furthermore, subjects who do progress to diabetes almost always have normal or impaired fasting glucose when their 2-hour value on the oral glucose tolerance test is >200 mg/dL,[45] and they do not have abnormal hemoglobin A1c until their fasting glucose is diagnostic of diabetes. Because this pattern is uncommon in the progression to clinical diabetes among those with T2DM, it may teach us about the relative contribution of impaired secretion and action to glucose tolerance status. Interestingly, although diminished insulin secretion is the sine qua non of autoimmune diabetes, little additional change in beta cell function can be detected in the year preceding diagnosis.[46] This suggests that other factors such as insulin resistance may play a role in subjects reaching the tipping point of diagnostic hyperglycemia. Thus, secondary prevention trials are feasible, they have a compelling rationale in that delaying or preventing onset of clinical disease and close monitoring for progression are of clear benefit to individuals, and these studies provide critical data to exploit as new therapies are developed.

Tertiary Prevention

The third area for T1DM prevention is tertiary prevention. As noted previously, at the time of diagnosis, almost all individuals continue to make significant amounts of insulin. Tertiary prevention aims to preserve remaining beta cell function. Because preservation of beta cell function after diagnosis will not cure the disease, it is worthwhile to consider why such studies are important. Data supporting such studies come from a cross-sectional analysis of individuals with and without residual beta cell function as well as data obtained from a retrospective analysis of the Diabetes Control and Complications Trial (DCCT) study. The landmark DCCT trial demonstrated that intensive therapy diminished both microvascular disease (retinopathy and proteinuria)[47] and macrovascular disease (cardiovascular endpoints).[48] However, these benefits were achieved at the risk of severe hypoglycemia, illustrating the Faustian inverse relationship between the risk and benefits of tight diabetes control.[47] In the presence of residual beta cell function, this relationship no longer exists.[49] Thus, a retrospective analysis of DCCT data found that those who had residual beta cell function within the intensively treated group had markedly less hypoglycemia and fewer complications that those without residual secretion. A C-peptide value greater than 0.2 pmol/mL was found to be clinically significant. Other smaller studies have demonstrated the same thing, that residual beta cell function reduces hypoglycemia and complications.[50] Although there remain important questions to answer (is there more benefit for C-peptide levels >0.2 pmol/mL, and how long must C-peptide be preserved to provide the observed benefits?), these data provide a compelling rationale for tertiary prevention studies to determine if therapy could prolong insulin secretion.

In addition to this rationale, it is generally felt that the immune-mediated beta cell destructive process is the same before and after diagnosis. Thus, a therapy that blocks beta cell destruction in tertiary prevention will likely be effective prior to disease onset and thus prevent clinical disease. Because thousands of individuals are diagnosed yearly, the research community can conduct multiple trials in this population, more rapidly determining the risks and benefits of a particular therapy than is possible in either primary or secondary prevention. These trials present their own challenges, including identifying and enrolling subjects soon after a diagnosis fraught with physical and emotional stress. However, because a key element of well-conducted tertiary prevention studies requires intensive diabetes control, all subjects benefit from the close attention and additional resources that come from participation in a clinical trial.

AUTOIMMUNITY AND IMMUNE RESPONSE

The role of the immune system is to maintain the delicate balance between effective responses against harmful pathogens and malignant cells and tolerance to self. Most self-reactive T cells are removed or become quiescent either centrally in the thymus or peripherally as mature T cells. Animal studies have suggested that residual self-reactive T cells can remain dormant because of low avidities of their T cell receptors (TCRs) for self-antigens, lack of costimulation from antigen-presenting cells (APCs), seclusion of self-antigens, or down-regulation of the effector T cells through regulatory T cell activity.[51] Which, if any, of these steps breaks down in human type 1 diabetes is not known. Furthermore, although the self-reactive effector T cells likely cause beta cell death, it is felt that the B cell arm of the immune system must also be engaged to propagate the deleterious response. It is likely that the controls on the immune response have to fail in multiple steps to result in autoimmune disease and that there is no one unique cause of autoimmunity.

In the face of imperfect knowledge of human disease, the choice of intervention generally relies on data from animal models of diabetes and/or data from other human autoimmune diseases. To understand these options, a simplistic concept of the immune response is illustrated in Figure 2. In general, a foreign (or, in the case of autoimmunity, self) protein or antigen is processed by an APC, which then presents the antigen in the context of the major histocompatibility complex to a T cell by interacting with a specific TCR, which includes the CD3 surface protein among others. This initial signal is then followed by secondary interactions between the APC and T cell. These costimulatory molecular pairs include the CD28 and CD40 ligands on the T cell interacting with CD80/86 and CD40 on the APC, respectively. Both are required for the production of an effective immune response; in the absence of costimulation, TCR signaling alone results in anergy or a lack of responsiveness to the signal. Intracellular pathways then respond, eventually propagating the immune attack through the release of cytokines and rapid antigen-specific T cell expansion. As this T cell effector part of the response resolves, a regulatory T cell response ensues, shutting down the immune attack.

From this description, it is apparent that multiple therapeutic targets are possible. One could affect the antigen itself, so that interaction with the immune response would result in an anergic or tolergenic signal. Therapies that alter the first or second signals between the APC and T cell or that have an impact on APC or T cell survival or proliferation could also disrupt the immune attack. Furthermore, manipulation of cytokine responses or an alteration of the balance between T effector and T regulatory cells would be expected to have efficacy in reducing autoimmunity. A full description of the myriad of therapies under development is beyond the scope of this review. Instead, we provide a brief overview of the therapies in recent, ongoing, or soon to begin clinical trials.

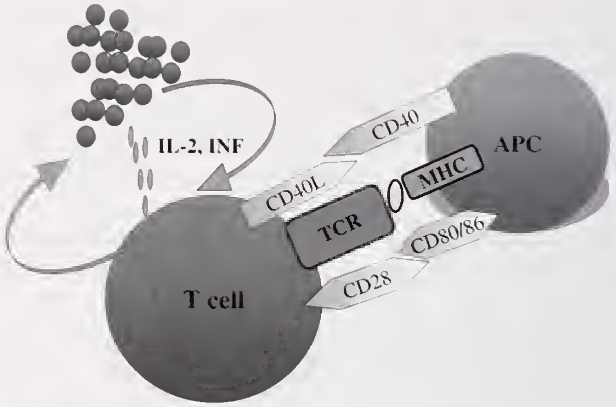


Fig 2. Cartoon of the immune response. **Abbreviations:** APC, antigen-presenting cell; IFNγ, interferon γ; IL-2, interleukin-2; MHC, major histocompatibility complex; TCR, T cell receptor. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Antigen-Specific Therapies

Insulin

In prospective studies of offspring of mothers with type 1 diabetes, anti-insulin antibodies are usually the first autoantibodies detected,[52] and in animal models and human T cell response to insulin and proinsulin, peptides have been reported.[53–56] These studies emphasize the importance of these antigens in the disease process and underlie the rationale for insulin-based clinical trials to interrupt beta cell immune destruction. DPT-1 tested the hypothesis that the administration of small doses of subcutaneous insulin in conjunction with a once yearly 4-day infusion of intravenous insulin would alter the natural history of disease in those at high risk of progression (>50% over 5 years). The hypothesis was that this early insulin treatment would act as an antigen to dampen the immune response and/or that it would induce beta cell rest; both had been demonstrated in the animal models to have efficacy in stopping disease. Unfortunately, this parenteral arm of DPT-1 did not prevent or delay the onset of disease,[42] possibly because the dose of insulin used was limited to alleviate potential concerns about hypoglycemia.

A second trial involving insulin was conducted by DPT-1 and included those relatives at intermediate risk of disease (ICA- and IAA-positive relatives with 25%–50% 5-year risk). This study was based on the concept of oral tolerance. The theory of oral tolerance is based on the understanding that the gut immune system serves a different purpose than the immune system elsewhere. The gut immune system is primarily designed to diminish the immune response. Anyone who has watched a toddler surreptitiously put debris into his mouth can understand this teleologically. If the gut's immune system responded to everything that it encountered, everyone would be continually allergic. Elegant animal studies have suggested that responses to antigens in the gut involve antigen-specific T cells that release regulatory cytokines.[57–59] Thus, delivery of insulin to the gut results in a breakdown into peptide fragments, which, while metabolically inactive, are recognized by the immune system. Insulin peptide-specific T cells with regulatory characteristics then travel to the islet, where they dampen the ongoing beta cell attack, thus inducing tolerance.

In the primary analysis of relatives selected and randomized in this DPT trial on the basis of ICA positivity with IAA $\geqslant$ 40 nU/mL, there were no untoward side effects from therapy, and unfortunately, oral insulin did not delay or prevent the development of diabetes.[43] Excitingly, in a post hoc analysis, there was evidence of heterogeneity of effect within the study cohort according to the level of IAA. The subgroup with confirmed IAA $\geqslant$ 80 nU/mL not only progressed to diabetes at a faster rate than those subjects who did not have confirmed IAA $\geqslant$ 80 nU/mL but also showed a potential beneficial effect of oral insulin: a delay of diabetes of more than 4 years! Such a result was clearly not only statistically significant ($P = 0.015$) but also of marked clinical significance as one can easily visualize by imagining a teen whose diagnosis is delayed until young adulthood. However, as this was a post hoc analysis, this result can be deemed only to be hypothesis-generating. This is the rationale underlying the new study testing whether oral insulin can delay or prevent diabetes. In this ongoing TrialNet oral insulin study, antibody-positive relatives with the appropriate risk profile are randomly assigned to once daily oral insulin capsules or placebo. Recruitment for this secondary prevention study is ongoing.

A pilot study is under development using oral insulin in a primary prevention trial (PrePoint Oral/Intranasal Insulin Trial). This study will enroll genetically at-risk infants to determine if oral insulin could prevent the development of autoantibodies and thus diabetes onset.

The Finnish Diabetes Prediction and Prevention Study (DIPP) also tested whether antigen-based insulin therapy would result in secondary prevention. In this trial, antibody-positive young children received nasal insulin or placebo, but unfortunately this therapy also did not delay or prevent disease in these very high risk subjects.[60] Study details have not yet been published, but it is likely that the dose used may have been a factor in this trial.

A pilot study using nasal insulin in older antibody-positive relatives was reported by Harrison *et al.*[53] to show a skewing of the immune response consistent with tolerance induction. A full trial to test whether this will change the course of the disease is underway in Australia and New Zealand (Intranasal Insulin Trial II) (INITT II) and is sponsored by the Diabetes Vaccine Development Center (DVDC).

Glutamic Acid Decarboxylase

Although its exact role in the pathogenesis of diabetes is unclear, an important autoantigen against which antibodies are detected in about 70% of patients with type 1 diabetes at the time of diagnosis is the enzyme glutamic acid decarboxylase.[61] Glutamic acid decarboxylase is present in the islets as well as the central nervous system and testes.[5]

Diamyd Corp. has developed a vaccine consisting of a highly purified and unmodified form of recombinant glutamic acid decarboxylase 65 formulated in aluminum hydroxide (alum). As with all antigen-based therapy, the goal is to alter the immune system's recognition and response to the antigen and induce a state of tolerance to further assault. In this preparation, the alum serves as an adjuvant to induce a protective immune response. In a tertiary prevention study in adults with so-called latent autoimmune diabetes of adults, administration of this subcutaneous vaccine was safe.[62,63] In unpublished data from a study in youth with T1DM, 2 doses of this vaccine a few months apart appeared to preserve beta cell function in comparison with those treated with placebo. The effect was particularly pronounced in those entering the study soon after diagnosis. TrialNet will soon be starting a clinical trial using 3 doses of subcutaneous recombinant glutamic acid decarboxylase 65 in alum in a double-masked tertiary prevention trial. Individuals within 3 months of diagnosis will be eligible to participate.

DVDC is also sponsoring a phase I trial testing the safety of administration of unique peptides from additional islet autoantigens. They were selected after elegant studies identifying these antigens as being naturally processed and presented by the body.[64] These dose escalation safety studies are ongoing in the United Kingdom (http://www.dvdc.org.au/index.php).

Immunosuppressant/Immunomodulatory Therapies

Thymoglobulin

Thymoglobulin is a Food and Drug Administration (FDA)–approved global immunosuppressant drug. Possible mechanisms of action include T cell clearance from the circulation and suppression of T cell activation, homing, and cytotoxic activities. It therefore blocks the T cell/APC activation at multiple points. Clinically used as induction therapy in organ transplantation, short-term treatment is designed to rapidly decrease autoreactive T cells. The Immune Tolerance Network (ITN) is sponsoring a phase II tertiary prevention trial in recently diagnosed individuals to determine if such aggressive therapy given over only a short period of time will effectively decrease the immune assault on the beta cells without unexpected side effects. This trial is ongoing and includes subjects within 6 weeks from the onset of their diabetes.

Mycophenolate Mofetil

Mycophenolate mofetil (MMF) is an FDA-approved immunosuppressant drug in clinical use to control autoimmune disease (eg, rheumatic disease) and to prevent rejection in organ transplantation. This oral medication inhibits the de novo pathway of guanosine nucleotide synthesis. Although other cell types can use a salvage pathway, T and B cells rely on the de novo pathway. Thus, MMF blocks T and B cell proliferation. Mild side effects such as gastrointestinal upset and diarrhea may be seen upon initiation of treatment, but prolonged therapy is well tolerated. TrialNet has recently finished recruitment for a 3-arm, placebo-controlled tertiary prevention trial using MMF as well as daclizumab (DZB). The concept is that short-term treatment (2 years of therapy) with MMF would sufficiently quiet the destructive immune response and allow the normal balance between effector and regulatory cells to be restored.

Daclizumab

DZB is a monoclonal antibody that inhibits interleukin-2 (IL-2)–mediated activation of lymphocytes. The IL-2 receptor consist of 3 components, α, β, and γ chains, and it appears on the cell surface of activated T cells. Upon engagement with IL-2, binding increases T cell growth and proliferation. DZB binds to the α chain of the IL-2 receptor, thus blocking this response. DZB is an intravenous drug FDA-approved for the prophylaxis and treatment of kidney transplant rejection. For the TrialNet tertiary prevention study, DZB/placebo was administered on 2 occasions 2 weeks apart. Well tolerated, this drug is expected to augment the effectiveness of MMF alone. Results of this multicenter TrialNet (TN) study are expected in 2009.

A phase I trial in adults with diabetes who still have residual beta cell function is testing the safety of the combination of IL-2 and rapamycin therapy under the auspices of the ITN. Although both IL-2 and rapamycin are FDA-approved drugs, they have never been coadministered. As noted previously, IL-2 serves to increase the growth and proliferation of all lymphocytes. This includes effector and regulatory cells. New knowledge about differences in the intracellular signaling pathways between effector and regulatory cells led to the observation that the immunosuppressant drug rapamycin would inhibit growth only of the effector cells.[65,66] Thus, the rationale behind the combination therapy is that IL-2 will drive both effector and regulatory cells, whereas rapamycin will keep effector cells in check, effectively increasing activated regulatory cells and thus inducing a tolerant state. Rapamycin is an orally administered, FDA-approved immunosuppressant drug widely used in organ transplantation. For this study, it is given for 3 months. IL-2 is administered as a subcutaneous injection 3 times a week for a month. Impressive results in animal models with this combination[67] support this phase I trial. Results should be available in 2009.

Anti-CD3 Monoclonal Antibodies

The key role of CD3 in the TCR complex was recognized decades ago and led to the development and use of an antibody designed to target and then destroy activated T cells with CD3 on the cell surface. Although effective in organ transplantation, marked side effects limited its use. Two new antibodies to the CD3 molecule were then developed, both designed to bind CD3 on activated T cells and modulate the response without causing massive T cell death, which was thought to underlie the side effects seen with earlier versions. Importantly, animal studies indicated that these new anti-CD3 drugs seemed to have a prolonged effect after short administration, thus inducing a tolerant state.[68] Several trials involving the treatment of subjects with new-onset T1DM have been conducted with a modified anti-CD3 monoclonal antibody. In an initial ITN-sponsored phase I trial, hOKT3γ1 (ala-ala) (now being developed as Teplizumab by Macrogenics) was administered over a 14-day course to recently diagnosed individuals between the ages of 7 and

30 in a randomized but not placebo-controlled trial. After 1 year of follow-up, C-peptide levels from the 4-hour mixed meal tolerance test were improved in 71% of treated subjects ($n = 21$) versus 21% of the control subjects ($n = 19$). Although there remained a difference between the groups with respect to C-peptide levels after 2 years, analysis of the follow-up data revealed a delayed but parallel decline in C-peptide response in comparison with controls.[69,70] This observation underlined the rationale for an ongoing follow-up ITN-sponsored trial (Autoimmunity-Blocking Antibody Therapy for Tolerance in Type 1 Diabetes) (ABATE), in which individuals will receive 2 courses of anti-CD3 treatment a year apart to assess whether the effect on beta cell function can be prolonged. Subjects 8 to 30 years old and within 6 weeks of diagnosis are eligible to enroll. The second modified anti-CD3 molecule, ChAglyCD3 (called TRX4 and under clinical development by Tolerx Co.), was studied as tertiary prevention in adults recently diagnosed with T1DM. Beta cell function in the treated groups was greater than that in controls, and there was a decrease in insulin requirement without differences in diabetes control as measured by hemoglobin A1c.[71] Side effects of both new anti-CD3 monoclonal antibodies, such as mild aches, fevers, and rashes, are likely attributable to cytokine release and are managed with anti-inflammatory and antihistamine therapy. These drugs are still in the early stages of drug development. Neither of these new antibodies is FDA-approved, although both continue in clinical trials.

The results from the tertiary prevention trials have led to the development of a TrialNet study in secondary prevention. In this study, antibody-positive relatives at very high risk of developing diabetes (ie, those with confirmed abnormal glucose tolerance who have about an 80% risk of disease) will be randomly assigned to treatment or not to determine if therapy can delay or prevent disease. Adults will receive placebo infusions, whereas children ages 8 to 18 will be randomized to treatment or close observation. Because the very high risk relatives are routinely being identified through the ongoing TrialNet natural history study, it is anticipated that enrollment will begin in 2008 and last about 2 years.

CTLA4-Ig

CTLA4-Ig (Abatacept) is an FDA-approved drug clinically used in autoimmune disease, including rheumatoid arthritis, and being tested in inflammatory bowel disease, multiple sclerosis, and lupus. This drug blocks costimulation between CD28 on the activated T cell and the APC, thus limiting the immune response. Clinical trials have documented the effectiveness and safety profile in very young children treated for juvenile rheumatoid arthritis. The drug has been administered as a monthly infusion lasting about 30 minutes, and there have been limited side effects associated with the infusion. A double-masked, placebo-controlled tertiary prevention trial of CTLA4-Ig in individuals ages 6 to 45 has just been launched by TrialNet. This study includes subjects within 3 months from their date of diabetes diagnosis with the aim of preserving residual beta cell function.

Rituximab

Because convincing evidence in animal models points to the role of T cells in destroying islet beta cells and others in autoimmune disease, it was somewhat surprising to observe that rituximab, a monoclonal antibody that destroys the immune system's B cells, was effective in patients with autoimmune disease. Originally developed and FDA-approved for the treatment of B cell lymphoma, it is now FDA-approved for the treatment of rheumatoid arthritis. This therapy is being tested in a double-masked, placebo-controlled tertiary prevention trial in recently diagnosed T1DM patients under the auspices of TrialNet. All subjects have been recruited for this trial, and preliminary information on whether this therapy preserves beta cell function is expected by the end of 2008.

Anakinra

Anakinra blocks the biological activity of IL-1 by competitively inhibiting IL-1 binding to the IL-1 type I receptor, which is expressed in a wide variety of tissues and organs. Depending on the concentration, this cytokine is known to be deleterious to cultured islet cells.[72] It was recently tested in individuals with T2DM. After 13 weeks of therapy, patients had improved beta cell function and improved glycemic control.[73] Anakinra will be used in a double-masked tertiary prevention trial sponsored by the Juvenile Diabetes Research Foundation (JDRF) in recently diagnosed subjects in Europe in 2008.

Other Approaches

Metabolic Control

The aforementioned studies are all primarily focused on interfering with the immune system. However, other factors clearly play a role in the development and clinical course. For example, the notion that residual beta cells would better be able to fight off immune attack if they were rested was one of the hypotheses that underlined the DPT-1 high-risk prevention trial discussed previously. Considerable data emphasize that hyperglycemia itself is deleterious to beta cells, either directly or indirectly. For example, in the DCCT, subjects in the intensive insulin therapy group maintained better beta cell function.[74] This has led to the planning of a new tertiary prevention trial using cutting-edge technology to maintain strict glucose control. Subjects within a week of diagnosis will be randomly assigned to usual care or to a closed-loop glucose-measuring, insulin-delivery system as inpatients to be followed by real-time continuous glucose monitoring to optimize insulin delivery. This TrialNet-sponsored study of intensive metabolic control should begin in 2008.

Beta Cell Growth

Obviously an intervention that diminishes or blocks the immune assault combined with a therapy that may augment islet beta cell growth and development would be of great interest. Despite a large number of animal and in vitro studies suggesting that glucagon-like peptide 1 mimetics such as exenatide have salutatory effects on inhibiting beta cell apoptosis or increasing beta cell mass,[75] it is not clear whether these effects occur in humans. Furthermore, although animal studies also suggest that epidermal growth factor and gastrin may promote beta cell mass,[76,77] other data suggest that such therapy would not likely be beneficial in humans.[78] Nonetheless, it is attractive to test the effects of these clinically available therapies. As such, TrialNet is considering combination trials with anti-CD3 and exenatide and/or antigen therapy in both secondary and tertiary prevention.

COORDINATED EFFORTS TOWARD THE PREVENTION OF TYPE 1 DIABETES

Multicenter Networks

Clinical trials to stop immune-mediated beta cell destruction are not new (see the reviews in refs. 79–82). Indeed, the cyclosporine trials demonstrated that immunotherapy can have an impact on the disease process.[83–85] However, most investigators feel that this is indeed a new era. The explosion of knowledge about fundamental aspects of the immune response, the plethora of therapies available to be tested, our increased sophistication about clinical trial design, and the increasing willingness of the diabetes community to participate in such studies underline the enthusiasm for such studies by funding agencies and pharmaceutical companies.

Although important studies are being conducted by individual researchers and groups around the world, the National Institutes of Health is funding 2 clinical trial networks to push the field forward. One, the ITN, is designed to conduct studies across autoimmunity, transplantation, and allergy with the aim of inducing tolerance. On a practical level, this implies using short-term therapy for long-term benefit. T1DM is one of the diseases in the ITN portfolio, and this network is currently conducting several phase I/phase II trials. More information about clinical trials underway through ITN can be found at www.immunetolerance.org.

Diabetes TrialNet is a network of clinical centers working in cooperation with affiliated sites throughout the United States, Canada, Finland, the United Kingdom, Italy, Germany, Australia, and New Zealand. This network, sponsored by the National Institutes of Health, JDRF, and American Diabetes Association, is designed to conduct multiple clinical trials in primary, secondary, and tertiary prevention. TrialNet also conducts other studies with the aim of facilitating standardization of clinical trial endpoints and testing of new assays that may one day serve as surrogate markers of the disease process or treatment effects. Updated information about TrialNet-sponsored research studies can be found at www.diabetestrialnet.org.

The JDRF, in conjunction with the Australian government, also funds another network, the DVDC. The DVDC conducts clinical trials to interrupt beta cell destruction. Information on the DVDC can be found at www.dvdc.org.au.

There are several significant advantages of multicenter networks. First, there can be consistencies in study design that allow for informal comparisons between trials. Second, the networks make primary and secondary prevention studies feasible and tertiary studies more rapidly accomplished. Third, publicly funded multicenter trials undergo a rigorous review before they begin and are carefully regulated. For example, a proposal submitted to TrialNet is reviewed for scientific rigor, ethical justification, clinical feasibility, and prioritization before it moves forward for a steering committee vote. An external advisory committee and independent data safety review board then evaluate the protocols, which are all operated under FDA regulations. No study moves forward without ethical or institutional review board approvals at each participating site. Safety and safety monitoring are important components of all of these studies. This is accomplished through an evaluation of clinical and laboratory data by the TrialNet medical monitor and a TrialNet safety committee. For studies involving immunotherapy, infectious disease consultants are readily available as needed, and regular reviews of all data are conducted by an independent data safety monitoring board. Similar structures are in place for other multicenter studies as well. Multicenter studies also provide widespread opportunity for participation by families.

Role of the Diabetes Clinical Team

Clinical investigators in T1DM need the help of the diabetes clinical team to let families and individuals with T1DM know that there are opportunities for participation in primary, secondary, and tertiary prevention trials. It is well recognized and understood that many families will decide not to participate in clinical research. However, too many families still are not informed about studies. Many later express frustration that they are outside the window of time from diagnosis and thus may no longer have a choice about participating or that they find out about screening programs for family members only after their second child is diagnosed with diabetes.

During this era of multiple clinical prevention trials, it is too much to ask busy clinicians to keep up with the myriad of approaches being tested. Indeed, this review does not even include all such studies! Rather, the clinician should understand the rationale underlying primary, secondary, and tertiary prevention studies and take an active role in letting patients and their families know about the studies. It may also be necessary to remind families and patients with diabetes that the knowledge gained from these prevention trials may be applied in the future when blocking of autoimmunity is needed for any biological beta cell replacement therapy. The clinician should refer families to the web, let them know about toll-free phone numbers, and encourage them to ask questions.

With your help, we hope to soon close the era of clinical trials and bring about an entirely new treatment paradigm for individuals at risk for or with diabetes.

DISCLOSURES

Potential conflict of interest: Dr. Greenbaum is Vice-Chair of Diabetes TrialNet, a member of the scientific review for the Immune Tolerance Network, a member of the board of directors of the Diabetes Vaccine Development Center, and is an investigator for clinical trials sponsored by these NIH or non-profit groups.

REFERENCES

1. Dittrich HM. History of the discovery of pancreatic diabetes by von Mering and Minkowski 1889. A historical overview on the occasion of the 100th anniversary [in German]. *Z Gesamte Inn Med* 1989; 44: 335–340.
2. Bottazzo GF, Florin-Christensen A, Doniach D. Islet-cell antibodies in diabetes mellitus with autoimmune polyendocrine deficiencies. *Lancet* 1974; 2: 1279–1283.
3. MacCuish A, Barnes E, Irvine W, Duncan L. Antibodies to islet cells in insulin-dependent diabetics with coexistent autoimmune disease. *Lancet* 1974; 2: 1529–1531.
4. Palmer J, Asplin C, Clemons P. Insulin autoantibodies in insulin-dependent diabetes before insulin treatment. *Science* 1983; 222: 1337–1339.
5. Baekkeskov S, Aanstoot H, Christgau S, et al. Identification of the 64kD autoantigen in insulin-dependent diabetes as the GABA-synthesizing enzyme glutamic acid decarboxylase. *Nature* 1990; 347: 151–156.
6. Lu J, Li Q, Xie H, et al. Identification of a second transmembrane protein tyrosine phosphatase, IA-2beta, as an autoantigen in insulin-dependent diabetes mellitus: precursor of the 37-kDa tryptic fragment. *Proc Natl Acad Sci U S A* 1996; 93: 2307–2311.
7. Wenzlau JM, Juhl K, Yu L, et al. The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. *Proc Natl Acad Sci U S A* 2007; 104: 17 040–17 045.
8. Foulis AK. The pathology of the endocrine pancreas in type 1 (insulin-dependent) diabetes mellitus. *Apmis* 1996; 104: 161–167.
9. Gepts W. Pathology and anatomy of the pancreas in juvenile diabetes mellitus. *Diabetes* 1965; 14: 619–633.
10. Kim SJ, Holbeck SL, Nisperos B, et al. Identification of a polymorphic variant associated with HLA-DQw3 and characterized by specific restriction sites within the DQ beta-chain gene. *Proc Natl Acad Sci U S A* 1985; 82: 8139–8143.
11. Menser MA, Forrest JM, Honeyman MC, Burgess JA. Letter: diabetes, HL-A antigens, and congenital rubella. *Lancet* 1974; 2: 1508–1509.
12. Nerup J, Platz P, Andersen OO, et al. HL-A antigens and diabetes mellitus. *Lancet* 1974; 2: 864–866.
13. Ganda OP, Srikanta S, Brink SJ, et al. Differential sensitivity to beta-cell secretagogues in "early," type I diabetes mellitus. *Diabetes* 1984; 33: 516–521.
14. Bleich D, Jackson RA, Soeldner JS, Eisenbarth GS. Analysis of metabolic progression to type I diabetes in ICA+ relatives of patients with type I diabetes. *Diabetes Care* 1990; 13: 111–118.
15. Srikanta A, Ganda OP, Gleason RE, et al. Pre-type I diabetes: linear loss of beta cell response to intravenous glucose. *Diabetes* 1984; 33: 717–720.
16. Faber O. Beta-cell function and diabetic control in insulin dependent diabetes mellitus. *Acta Endocrinol Suppl (Copenh)* 1985; 272: 73–77.
17. Hoogwerf BJ, Rich SS, Barbosa JJ. Meal-stimulated C-peptide and insulin antibodies in type I diabetic

subjects and their nondiabetic siblings characterized by HLA-DR antigens. *Diabetes* 1985; 34: 440–445.

18. Scholin A, Bjorklund L, Borg H, et al. Islet antibodies and remaining beta-cell function 8 years after diagnosis of diabetes in young adults: a prospective follow-up of the nationwide Diabetes Incidence Study in Sweden. *J Intern Med* 2004; 255: 384–391.
19. Incidence and trends of childhood type 1 diabetes worldwide 1990–1999. *Diabet Med* 2006; 23: 857–866.
20. Harjutsalo V, Podar T, Tuomilehto J. Cumulative incidence of type 1 diabetes in 10,168 siblings of Finnish young-onset type 1 diabetic patients. *Diabetes* 2005; 54: 563–569.
21. Karvonen M, Viik-Kajander M, Moltchanova E, et al. Incidence of childhood type 1 diabetes worldwide. Diabetes Mondiale (DiaMond) Project Group. *Diabetes Care* 2000; 23: 1516–1526.
22. Lammi N, Taskinen O, Moltchanova E, et al. A high incidence of type 1 diabetes and an alarming increase in the incidence of type 2 diabetes among young adults in Finland between 1992 and 1996. *Diabetologia* 2007; 50: 1393–1400.
23. Passa P. Diabetes trends in Europe. *Diabetes Metab Res Rev* 2002; 18(suppl 3): S3–S8.
24. TEDDY Study Group. The Environmental Determinants of Diabetes in the Young (TEDDY) study: study design. *Pediatr Diabetes* 2007; 8: 286–298.
25. Hermann R, Knip M, Veijola R, et al. Temporal changes in the frequencies of HLA genotypes in patients with type 1 diabetes–indication of an increased environmental pressure? *Diabetologia* 2003; 46: 420–425.
26. Gillespie KM, Bain SC, Barnett AH, et al. The rising incidence of childhood type 1 diabetes and reduced contribution of high-risk HLA haplotypes. *Lancet* 2004; 364: 1699–1700.
27. Barker JM, Barriga KJ, Yu L, et al. Prediction of autoantibody positivity and progression to type 1 diabetes: Diabetes Autoimmunity Study in the Young (DAISY). *J Clin Endocrinol Metab* 2004; 89: 3896–3902.
28. Kimpimaki T, Kupila A, Hamalainen AM, et al. The first signs of beta-cell autoimmunity appear in infancy in genetically susceptible children from the general population: the Finnish Type 1 Diabetes Prediction and Prevention Study. *J Clin Endocrinol Metab* 2001; 86: 4782–4788.
29. Schenker M, Hummel M, Ferber K, et al. Early expression and high prevalence of islet autoantibodies for DR3/4 heterozygous and DR4/4 homozygous offspring of parents with type I diabetes: the German BABYDIAB study. *Diabetologia* 1999; 42: 671–677.
30. Virtanen SM, Saukkonen T, Savilahti E, et al. Diet, cow's milk protein antibodies and the risk of IDDM in Finnish children. Childhood Diabetes in Finland Study Group. *Diabetologia* 1994; 37: 381–387.
31. Norris JM, Beaty B, Klingensmith G, et al. Lack of association between early exposure to cow's milk protein and beta-cell autoimmunity. Diabetes Autoimmunity Study in the Young (DAISY). *JAMA* 1996; 276: 609–614.
32. TRIGR Study Group. Study design of the Trial to Reduce IDDM in the Genetically at Risk (TRIGR). *Pediatr Diabetes* 2007; 8: 117–137.
33. Norris JM, Barriga K, Klingensmith G, et al. Timing of initial cereal exposure in infancy and risk of islet autoimmunity. *JAMA* 2003; 290: 1713–1720.
34. Ziegler AG, Schmid S, Huber D, et al. Early infant feeding and risk of developing type 1 diabetes-associated autoantibodies. *JAMA* 2003; 290: 1721–1728.
35. Fuchtenbusch M, Ziegler AG, Hummel M. Elimination of dietary gluten and development of type 1 diabetes in high risk subjects. *Rev Diabet Stud* 2004; 1: 39–41.
36. Simopoulos AP. Essential fatty acids in health and chronic disease. *Am J Clin Nutr* 1999; 70: 560S–569S.
37. Litherland SA, Xie XT, Hutson AD, et al. Aberrant prostaglandin synthase 2 expression defines an antigen-presenting cell defect for insulin-dependent diabetes mellitus. *J Clin Invest* 1999; 104: 515–523.
38. Stene LC, Joner G. Use of cod liver oil during the first year of life is associated with lower risk of childhood-onset type 1 diabetes: a large, population-based, case-control study. *Am J Clin Nutr* 2003; 78: 1128–1134.
39. Mathieu C, Badenhoop K. Vitamin D and type 1 diabetes mellitus: state of the art. *Trends Endocrinol Metab* 2005; 16: 261–266.
40. Adorini L. Intervention in autoimmunity: the potential of vitamin D receptor agonists. *Cell Immunol* 2005; 233: 115–124.
41. Wicklow BA, Taback SP. Feasibility of a type 1 diabetes primary prevention trial using 2000 IU vitamin D3 in infants from the general population with increased HLA-associated risk. *Ann N Y Acad Sci* 2006; 1079: 310–312.
42. Diabetes Prevention Trial Study Group. Effects of insulin in relatives of patients with type 1 diabetes mellitus. *N Engl J Med* 2002; 346: 1685–1691.
43. Skyler JS, Krischer JP, Wolfsdorf J, et al. Effects of oral insulin in relatives of patients with type 1 diabetes: the Diabetes Prevention Trial-Type 1. *Diabetes Care* 2005; 28: 1068–1076.
44. Gale EA, Bingley PJ, Emmett CL, Collier T. European Nicotinamide Diabetes Intervention Trial (ENDIT): a randomised controlled trial of intervention before the onset of type 1 diabetes. *Lancet* 2004; 363: 925–931.
45. Greenbaum CJ, Cuthbertson D, Krischer JP. Type 1 diabetes manifested solely by 2-h oral glucose tolerance test criteria. *Diabetes* 2001; 50: 470–476.
46. Sosenko JM, Palmer JP, Greenbaum CJ, et al. Patterns of metabolic progression to type 1 diabetes in the Diabetes Prevention Trial-Type 1. *Diabetes Care* 2006; 29: 643–649.
47. Diabetes Control and Complications Trial. *Henry Ford Hosp Med J* 1983; 31: 115–116.
48. Nathan DM, Lachin J, Cleary P, et al. Intensive diabetes therapy and carotid intima-media thickness in type I diabetes mellitus. *N Engl J Med* 2003; 348: 2294–2303.
49. Steffes MW, Sibley S, Jackson M, Thomas W. Beta-cell function and the development of diabetes-related complications in the diabetes control and complications trial. *Diabetes Care* 2003; 26: 832–836.
50. Sjoberg S, Gjotterberg M, Berglund L, et al. Residual C-peptide excretion is associated with a better long-term glycemic control and slower progress of retinopathy in type 1 (insulin-dependent) diabetes mellitus. *J Diabet Complications* 1991; 5: 18–22.
51. Shevach EM. Regulatory T cells in autoimmunity. *Annu Rev Immunol* 2000; 18: 423–449.
52. Ziegler AG, Hillebrand B, Rabl W, et al. On the appearance of islet associated autoimmunity in offspring of diabetic mothers: a prospective study from birth. *Diabetologia* 1993; 36: 402–408.

53. Harrison LC, Honeyman MC, Steele CE, et al. Pancreatic beta-cell function and immune responses to insulin after administration of intranasal insulin to humans at risk for type 1 diabetes. *Diabetes Care* 2004; 27: 2348–2355.
54. Harrison L, Chu X, DeAizpurua H, et al. Islet reactive T cells are a marker of preclinical insulin-dependent diabetes. *J Clin Invest* 1992; 89: 1161–1165.
55. Peakman M, Tree TI, Endl J, et al. Characterization of preparations of GAD65, proinsulin, and the islet tyrosine phosphatase IA-2 for use in detection of autoreactive T-cells in type 1 diabetes: report of phase II of the Second International Immunology of Diabetes Society Workshop for Standardization of T-cell assays in type 1 diabetes. *Diabetes* 2001; 50: 1749–1754.
56. Pinkse GG, Boitard C, Tree TI, et al. HLA class I epitope discovery in type I diabetes: independent and reproducible identification of proinsulin epitopes of CD8 T cells–report of the IDS T Cell Workshop Committee. *Ann N Y Acad Sci* 2006; 1079: 19–23.
57. Ke Y, Pearce K, Lake JP, et al. Gamma delta T lymphocytes regulate the induction and maintenance of oral tolerance. *J Immunol* 1997; 158: 3610–3618.
58. Weiner HL. Treatment of autoimmune diseases by oral tolerance to autoantigens. *Autoimmunity* 1993; 15: 6–7.
59. Weiner HL, Friedman A, Miller A, et al. Oral tolerance: immunologic mechanisms and treatment of animal and human organ-specific autoimmune diseases by oral administration of autoantigens. *Annu Rev Immunol* 1994; 12: 809–837.
60. Simell O. *Proc Immunol Diabetes Soc* 2007 unpublished report from Immunology of Diabetes Society meeting presentation.
61. Bingley PJ, Bonifacio E, Mueller PW. Diabetes Antibody Standardization Program: first assay proficiency evaluation. *Diabetes* 2003; 52: 1128–1136.
62. Agardh CD, Cilio CM, Lethagen A, et al. Clinical evidence for the safety of GAD65 immunomodulation in adult-onset autoimmune diabetes. *J Diabetes Complications* 2005; 19: 238–246.
63. Bekris LM, Jensen RA, Lagerquist E, et al. GAD65 autoantibody epitopes in adult patients with latent autoimmune diabetes following GAD65 vaccination. *Diabet Med* 2007; 24: 521–526.
64. Peakman M, Stevens EJ, Lohmann T, et al. Naturally processed and presented epitopes of the islet cell autoantigen IA-2 eluted from HLA-DR4. *J Clin Invest* 1999; 104: 1449–1457.
65. Bensinger SJ, Walsh PT, Zhang J, et al. Distinct IL-2 receptor signaling pattern in CD4+CD25+ regulatory T cells. *J Immunol* 2004; 172: 5287–5296.
66. Tian L, Lu L, Yuan Z, et al. Acceleration of apoptosis in CD4+CD8+ thymocytes by rapamycin accompanied by increased CD4+CD25+ T cells in the periphery. *Transplantation* 2004; 77: 183–189.
67. Rabinovitch A, Suarez-Pinzon WL, Shapiro AM, et al. Combination therapy with sirolimus and interleukin-2 prevents spontaneous and recurrent autoimmune diabetes in NOD mice. *Diabetes* 2002; 51: 638–645.
68. Chatenoud L, Thervet E, Primo J, Bach JF. Anti-CD3 antibody induces long-term remission of overt autoimmunity in nonobese diabetic mice. *Proc Natl Acad Sci U S A* 1994; 91: 123–127.
69. Herold KC, Gitelman SE, Masharani U, et al. A single course of anti-CD3 monoclonal antibody hOKT3gamma1(Ala-Ala) results in improvement in C-peptide responses and clinical parameters for at least 2 years after onset of type 1 diabetes. *Diabetes* 2005; 54: 1763–1769.
70. Herold KC, Hagopian W, Auger JA, et al. Anti-CD3 monoclonal antibody in new-onset type 1 diabetes mellitus. *N Engl J Med* 2002; 346: 1692–1698.
71. Keymeulen B, Vandemeulebroucke E, Ziegler AG, et al. Insulin needs after CD3-antibody therapy in new-onset type 1 diabetes. *N Engl J Med* 2005; 352: 2598–2608.
72. Mandrup-Poulsen T. The role of interleukin-1 in the pathogenesis of IDDM. *Diabetologia* 1996; 39: 1005–1029.
73. Larsen CM, Faulenbach M, Vaag A, et al. Interleukin-1-receptor antagonist in type 2 diabetes mellitus. *N Engl J Med* 2007; 356: 1517–1526.
74. The Diabetes Control and Complications Trial Research Group. Effect of intensive therapy on residual beta-cell function in patients with type 1 diabetes in the diabetes control and complications trial. A randomized, controlled trial. *Ann Intern Med* 1998; 128: 517–523.
75. Drucker DJ. Biologic actions and therapeutic potential of the proglucagon-derived peptides. *Nat Clin Pract Endocrinol Metab* 2005; 1: 22–31.
76. Brand SJ, Tagerud S, Lambert P, et al. Pharmacological treatment of chronic diabetes by stimulating pancreatic beta-cell regeneration with systemic co-administration of EGF and gastrin. *Pharmacol Toxicol* 2002; 91: 414–420.
77. Suarez-Pinzon WL, Yan Y, Power R, et al. Combination therapy with epidermal growth factor and gastrin increases beta-cell mass and reverses hyperglycemia in diabetic NOD mice. *Diabetes* 2005; 54: 2596–2601.
78. Meier JJ, Butler AE, Galasso R, et al. Increased islet beta cell replication adjacent to intrapancreatic gastrinomas in humans. *Diabetologia* 2006; 49: 2689–2696.
79. Greenbaum CJ. Type 1 diabetes intervention trials: what have we learned? A critical review of selected intervention trials. *Clin Immunol* 2002; 104: 97–104.
80. Pozzilli P. Immuno-intervention and preservation of beta-cell function in type 1 diabetes. *Diabetes Metab Res Rev* 2007; 23: 255–256.
81. Skyler JS. Prediction and prevention of type 1 diabetes: progress, problems, and prospects. *Clin Pharmacol Ther* 2007; 81: 768–771.
82. Staeva-Vieira T, Peakman M, von Herrath M. Translational mini-review series on type I diabetes: immune-based therapeutic approaches for type 1 diabetes. *Clin Exp Immunol* 2007; 148: 17–31.
83. Assan R, Feutren G, Sirmai J, et al. Plasma C-peptide levels and clinical remissions in recent-onset type I diabetic patients treated with cyclosporin A and insulin. *Diabetes* 1990; 39: 768–774.
84. Stiller CR, Laupacis A, Dupre J, et al. Cyclosporine for treatment of early type I diabetes: preliminary results. *N Engl J Med* 1983; 308: 1226–1227.
85. Stiller CR, Dupre J, Gent M, et al. Effects of cyclosporine immunosuppression in insulin-dependent diabetes mellitus of recent onset. *Science* 1984; 223: 1362–1367.

Mount Sinai Journal of Medicine 75:398–400, 2008

CONTROVERSY

Contemporary Issues in Management of Carotid Occlusive Disease

Gregg S. Landis, MD,[1] and Peter L. Faries, MD[2]

[1] Division of Vascular Surgery, New York Hospital Queens, Flushing, NY
[2] Division of Vascular Surgery, Mount Sinai Medical School, New York, NY

Atherosclerotic disease at the bifurcation of the carotid artery is a common cause of ischemic strokes, resulting in significant morbidity and mortality annually in the United States. Over 700,000 new strokes are diagnosed nationally each year. Moreover, nearly 150,000 patients die each year as a result of stroke.[1] It is estimated that 20% to 30% of strokes are caused by carotid artery disease.[2] Risk factors such as hypertension, hyperlipidemia, diabetes mellitus, and tobacco use contribute to the development of occlusive plaques at the origin of the internal carotid artery. Strokes are typically caused by an embolism of atheromatous debris from the carotid bulb and to a much lesser extent by hemodynamic decrements in flow through highly stenotic lesions. Since the 1950s, when surgical procedures were first developed to intervene in this process, stroke prevention for this disease process has become feasible.[3] Carotid endarterectomy is a surgical procedure in which the diseased intima and media are excised through a transcervical arteriotomy. Throughout the 1970s and 1980s, refinements in the technique of carotid endarterectomy allowed it to emerge as a safe, reliable, reproducible, and widespread procedure. Advances such as intraprocedural shunting of blood, electrophysiologic monitoring, and prosthetic patch angioplasty reduced both the short-term and long-term complications of the procedure. Postoperative myocardial infarction and intraprocedural stroke were 2 of the most troublesome complications affecting this procedure. By the 1990s, the incidence of cardiac complications was approximately 2%, and the procedural stroke rate was about 1% to 3%.[4]

As stroke prevention was the primary goal of therapy, proponents of endarterectomy needed to demonstrate improved and lasting efficacy over both medical therapy and the natural history of the disease process itself. This led to a series of large multicenter, prospective, and randomized trials comparing endarterectomy to medical therapy, which at that time was aspirin monotherapy. Published in 1991, the North American Symptomatic Carotid Endarterectomy Trial (NASCET) established the superior efficacy of surgical carotid endarterectomy over medical treatment in symptomatic patients with a high-grade stenosis.[5] Symptomatic patients are defined as those who have suffered from a transient ischemic attack, transient monocular blindness (amaurosis fugax), or a fixed stroke from an ipsilateral carotid lesion. NASCET data demonstrated a cumulative risk of ipsilateral stroke at 2 years of 26% in the medical arm of the study versus 9% in the surgical arm. This corresponded to an absolute risk reduction of 17% ± 3.5% for patients undergoing surgery. Several years later, the Asymptomatic Carotid Artery Stenosis Study (ACAS) established similar stroke risk reductions in asymptomatic patients undergoing surgical revascularization in comparison with medical therapy alone.[6] With a median follow-up of 2.7 years, the aggregate risk of ipsilateral stroke during a 5-year period was 11% in the medical cohort and 5.1% in the surgical cohort. This corresponded to an aggregate risk reduction of 53%. In these asymptomatic patients, the marginal benefit of the procedure was not realized when there were small increases in perioperative stroke rates. Therefore, only those surgeons with low rates of technical

Address Correspondence to:

Gregg S. Landis, MD
Division of Vascular Surgery
New York Hospital Queens
Flushing, NY
Email: grl9006@nyp.org

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20053

© 2008 Mount Sinai School of Medicine

complications could justify performing this procedure on asymptomatic patients.

Shortly after the publication of these studies, dramatic increases in the number of carotid endarterectomies were seen nationally. During the 5-year period following 1991, the rate of carotid endarterectomies rose 94% nationally.[7] Because of the stringent exclusion criteria used in both landmark studies, the applicability of these data to various subgroups such as women and medically high-risk individuals was scrutinized. What followed over the next decade were numerous smaller single-institution studies evaluating the performance of these procedures in patients with comorbidities such as unstable angina, atrial fibrillation, uncontrolled hypertension, and renal insufficiency. Importantly, these were factors used as exclusion criteria in NASCET and ACAS.[8] Generally, these subsequent data demonstrated the safety of carotid endarterectomy in patients characterized as high-risk. In fact, many authors have questioned whether such a category actually exists and have proposed that the vast majority of patients with significant cardiovascular, pulmonary, and renal disease can undergo carotid endarterectomy with acceptable risk.[9] The actual number of patients that cannot undergo carotid surgery without poor outcomes is likely to be much smaller than what is widely believed.

Carotid artery stenting is a minimally invasive percutaneous procedure in which a metallic stent is delivered transarterially under fluoroscopic guidance to the carotid bifurcation. The technique was first described by Mathias in 1980 using simple balloon angioplasty and then in 1989 using a metallic stent.[10] Initially, carotid stenting was fraught with a high incidence of intraprocedural strokes as a result of embolization from the friable plaque at the carotid bifurcation. With its prohibitive morbidity rate, carotid stenting would not be able to compete with the well-documented performance of surgery, and it seemed doomed to incidental usage. However, with the advent of distal protection devices designed to capture particulate matter released during manipulation of the lesion, safety coupled with a minimally invasive procedure became realizable. These protection devices are delivered first through the lesion and placed in the distal internal carotid artery. Once deployed, the majority of the devices act like a windsock or filter, capturing macroscopic particulate matter while permitting the flow of blood. The first major randomized, prospective trial to show results comparable to those of surgery was the 2004 Stenting and Angioplasty with Protection in Patients at High Risk for Endarterectomy (SAPPHIRE) trial.[11] Although statistically powered to demonstrate only noninferiority, this study opened the door to stenting as a viable alternative to carotid endarterectomy. On the basis of the findings of the SAPPHIRE trial, in 2007 the Centers for Medicare and Medicaid Services approved coverage for symptomatic high-risk patients undergoing carotid stenting.[12]

Currently, multiple investigations are underway that are attempting to define the role of carotid stenting with respect to surgery. Most trials to date have focused on evaluating the safety and efficacy of stenting in patients considered at high risk for surgery. These include patients with anatomic considerations such as prior neck irradiation, tracheostomy, a high cervical carotid bifurcation, or recurrent stenosis. In theory, these factors confound the ability of the surgeon to safely perform the procedure without local complications. Physiologic comorbidities are also considered, such as unstable angina, recent myocardial infarction, severe congestive heart failure, contralateral carotid occlusion, and advanced age. Obviously, these factors can precipitate perioperative cardiopulmonary or cerebrovascular events, yet substantial data, albeit from smaller single-center series, have demonstrated the ability to safely perform carotid endarterectomies under all of the aforementioned conditions.[9] Comparisons between carotid stenting and surgery in these high-risk patients are therefore difficult to interpret. The Carotid Revascularization Endarterectomy Versus Stenting Trial (CREST)[13] is the largest active randomized clinical trial currently enrolling patients at standard risk, and it should help to clarify some of these issues. Inclusion criteria for CREST are similar to those for NASCET and ACAS, and 2500 patients are expected to enroll by mid 2008. It is anticipated that CREST will yield definitive data regarding the performance of carotid stenting in real-world patients.

There are clearly subsets of patients that are at high risk for carotid stenting procedures. Unfavorable anatomy of the aortic arch and great vessels, such as calcification, tortuosity, and common carotid lesions, can increase the periprocedural complication rate.[14] These factors appear to coincide with increasing age. This may account for the increased complication rate, observed in some series, with carotid stenting in octogenarians.[15] Advances in device design and operator technical proficiency may ameliorate some of these issues.

With the addition of statins, beta-blockers, and newer antiplatelet medications such as clopidrogel to the medical armamentarium, the notion of "best medical therapy" has evolved. Plaque stabilization, inhibition of inflammation, and a variety of other proposed pleiotropic effects from statins have changed the natural history of atherosclerotic lesions.[16] The cardioprotective effects of sustained beta-blockade have also had a positive impact on lowering perioperative cardiac complications. These effects are likely to change the clinical course of carotid occlusive lesions and ultimately reduce the risk of nonsurgical management of this disease. As a result, comparisons of contemporary carotid stent performance, in which patients are treated with multimodal medical therapy, to historical data from NASCET and ACAS, in which aspirin was the sole medical treatment, are problematic. Moreover, the marginal incremental stroke reduction seen in asymptomatic patients undergoing surgery is small; therefore, the application of contemporary medical therapy may erase the narrow beneficial effect of surgery. Thus, the modern utility of carotid endarterectomy in asymptomatic patients needs to be revisited.

A newer generation of studies such the Transatlantic Asymptomatic Carotid Intervention Trial (TACIT)[17] may reconcile many of these issues. TACIT is a multicenter, prospective, 3-way randomization trial evaluating both high-risk and low-risk asymptomatic patients with greater than 70% stenosis. Patients will be randomized to the following 3 treatment arms: carotid endarterectomy, carotid stenting, and contemporary multimodal medical management. Studies such as this will be needed to determine the applicability of these treatments to various types of patients.

In conclusion, carotid endarterectomy remains a safe and effective treatment for carotid bifurcation disease in most patients. The notion that a substantial number of patients are at high risk for this surgical procedure is not supported by the literature, yet the justification for the adoption of carotid stenting has been predicated on the need for an alternative to surgery in high-risk patients. Clearly, carotid stenting is here to stay, and its precise role in the management of carotid disease remains to be determined. Lastly, major advances in medical therapy for atherosclerotic disease are likely to improve stroke-free survival in patients affected by carotid bifurcation disease. In order to address these issues, a series of contemporary randomized trials are needed. Until that time, careful patient selection and judicious use of new technologies are the best means of ensuring optimal patient outcomes.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. American Stroke Association. Impact of stroke. Available at: http://www.strokeassociation.org/presenter.jhtml?identifier=1033. Accessed June 2008.
2. Timsit SG, Sacco RL, Mohr JP, et al. Early clinical differentiation of cerebral infarction from severe atherosclerotic stenosis and cardioembolism. *Stroke* 1992; 23: 486–491.
3. Eastcott HHG, Pickering GW, Rob CG. Reconstruction of internal carotid artery in a patient with intermittent attacks of hemiplegia. *Lancet* 1954; 2: 994–996.
4. Sumner DS. Nonstroke complications of carotid endarterectomy. In: Bernhard VM, Towne JB, eds. *Complications in Vascular Surgery*. St. Louis, MO: Quality Medical Publishing; 1991; 471–500.
5. North American Symptomatic Carotid Endarterectomy Trial Collaborators. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade stenosis. *N Engl J Med* 1991; 325: 445–453.
6. Executive Committee for the Asymptomatic Carotid Atherosclerosis Study. Endarterectomy for asymptomatic carotid artery stenosis. *JAMA* 1995; 273: 1421–1428.
7. Gillum RF. Epidemiology of carotid endarterectomy and cerebral angiography in the United States. *Stroke* 1995; 26: 1724–1728.
8. Morasch MD, Parker MA, Feinglass J, et al. Carotid endarterectomy: characterization of recent increases in procedure rates. *J Vasc Surg* 2000; 31: 901–909.
9. Landis GS, Faries PL. A critical look at "high-risk" in choosing the proper intervention for patients with carotid bifurcation disease. *Semin Vasc Surg* 2007; 20: 199–204.
10. Mathias K. Historical background: 25 years of endovascular therapy for obstructive carotid artery disease. In: Al-Mubarak N, Roubin GS, Iyer SS, Vitek JJ, eds. *Carotid Artery Stenting*. Philadelphia, PA: Lippincott, Williams & Wilkins; 2004; 18–22.
11. Yadav JS, Wholey MH, Kuntz RE, et al. Protected carotid artery stenting versus endarterectomy in high-risk patients. *N Engl J Med* 2004; 351: 1493–1501.
12. Centers for Medicare and Medicaid Services. Decision memo for percutaneous transluminal angioplasty (PTA) of the carotid artery concurrent with stenting (CAG-00085R3). Available at: www.cms.hhs.gov/mcd/viewdecisionmemo.asp?id=194. Accessed June 2008.
13. Hobson RW II. Update on the Carotid Revascularization Endarterectomy Versus Stent Trial (CREST) protocol. *J Am Coll Surg* 2002; 194(suppl 1): S9–S14.
14. Lam RC, Lin SC, DeRubertis B, et al. The impact of increasing age on anatomic factors affecting carotid angioplasty and stenting. *J Vasc Surg* 2007; 45: 875–880.
15. Lin SC, Trocciola SM, Rhee J, et al. Analysis of anatomic factors and age in patients undergoing carotid angioplasty and stenting. *Ann Vasc Surg* 2005; 19: 798–804.
16. Rubba P. Effects of atorvastatin on the different phases of atherogenesis. *Drugs* 2007; 67(suppl 1): 17–27.
17. Katzen BT. The Transatlantic Asymptomatic Carotid Intervention Trial. *Endovasc Today* 2005; (September): 49–50.

EVIDENCE-BASED MEDICINE

EBM on call: Management of Minimally Symptomatic Inguinal Hernia

Deborah Korenstein and Thomas McGinn

Division of General Internal Medicine, Department of Medicine, Mount Sinai School of Medicine, New York, NY

This is the third in a series of articles in the *Mount Sinai Journal of Medicine* on evidence-based medicine; this series will emphasize applications of evidence-based medicine concepts and procedures to clinical practice. The articles will focus on the retrieval, interpretation, and application of evidence related to a clinical problem. The discussion will be based on the following case.

THE CASE

A 67-year-old male with a history of mild asthma presents to a general surgeon for evaluation of a right inguinal hernia. The patient is healthy; the hernia had been noticed by his primary care physician during a routine physical examination. The patient states that he may have noticed an occasional vague ache in the right groin for the last several months but that overall he has not had any bothersome symptoms. On physical examination, the surgeon finds an indirect inguinal hernia that is easily reducible. The remainder of the examination is normal.

The patient explains that he would prefer to avoid surgery if possible. He asks the surgeon if it is safe to watch the hernia and undergo surgery only if it becomes more symptomatic.

Address Correspondence to:

Deborah Korenstein
Division of General Internal Medicine
Department of Medicine
Mount Sinai School of Medicine
New York, NY
Email: deborah.korenstein@mountsinai.org

FOCUSED CLINICAL QUESTION

To determine if it is safe for this patient to wait for surgery, the surgeon decides to review the relevant medical literature. He begins by formulating a focused question consisting of 4 elements: the patient (P), the intervention (I), the comparison (C), and the outcome (O) referred to as a "PICO" question.[1] The explicit inclusion of these 4 components allows for a more efficient and directed search of the literature and helps to clarify the exact problem of interest. In this case, the question's components are as follows: in patients with minimally symptomatic inguinal hernias (P), does watchful waiting (I), in comparison with immediate surgical repair (C), lead to more complications or worse quality of life (O)?

SEARCH OF THE LITERATURE

The surgeon searches Medline using the PubMed interface to find his answer. He begins with a keyword search by typing his terms directly into the search box. He types in "inguinal hernia repair" and obtains 3370 citations. To focus his search on more relevant articles, he clicks the "Limits" button and selects "Human," "English," and "Randomized Controlled Trial" within the limits screen. Applying limits reduces the number of citations by including only those that meet certain criteria, such as a certain language, article type, or subject age. The search with limits applied yields 341 articles, which are still too many to review. The surgeon clicks the "Limits" button again to deselect the applied limits (note that once limits are chosen they will be applied to all subsequent searches in that session unless the "Limits" button or particular limits are deselected) and types "inguinal hernia repair waiting" into the search box. This search

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20061

© 2008 Mount Sinai School of Medicine

yields 34 citations. The surgeon clicks "Limits" again and selects "Randomized Controlled Trial"; this narrows the search to 5 citations. The fourth on the list is entitled "Watchful Waiting vs Repair of Inguinal Hernia in Minimally Symptomatic Men: A Randomized Controlled Trial" by Fitzgibbons *et al.*,[2] and it was published in *JAMA* in January 2006. The surgeon decides to review that article.

Validity Criteria for Treatment Trials

The study involved 720 men from North American centers with minimally symptomatic inguinal hernias, which were defined as hernias that were easily reducible and did not cause discomfort limiting usual activities. Patients were referred to the study by their primary care physician or surgeon or responded to advertisements for the study. The primary outcomes were pain or discomfort interfering with usual activities at 2 years' follow-up and a change from the baseline in the physical component score of the Short Form 36 Version 2 health-related quality-of-life survey. Secondary outcomes included complications, pain, functional status, activity levels, and patient satisfaction. The major issues of study design and validity are described below[4]:

- The study was **randomized**. Patients were randomly assigned to 2 groups, one of which received standard open surgical hernia repair and the other of which received watchful waiting. Patients in the watchful waiting group were given written instructions to follow their hernias and watch for changes in symptoms, and they were examined by study physicians 6 months into the study and annually thereafter. They were instructed to contact their doctor if they developed any problems during the intervening months
- **Allocation concealment** was performed adequately. Allocation concealment refers to a system established by the investigators to ensure that clinicians enrolling a potential trial participant cannot predict to which group that participant will be assigned once enrolled (reference consort statement web site).[3] This is important in that it eliminates the ability of physicians to have an impact on enrollment by accurately predicting the group to which their patient will be assigned. For example, in a randomized trial of a new medication versus placebo, a physician who knows that the next subject will be assigned to the placebo group might dissuade a sicker patient from enrolling in that trial. Similarly, a physician who knows that the next enrolled patient will receive the active drug might push for his or her sick patient to enroll. Allocation concealment differs from blinding in that it occurs before patients are enrolled in the study at all. There are many methods to achieve allocation concealment. Many involve the computer generation of group assignments or third parties such as central telephone centers or pharmacies. Local allocation concealment can be accomplished with sealed envelopes or containers, but it is more challenging. In the study by Fitzgibbons *et al.*,[2] allocation concealment was achieved through computer generation at a central site. Local investigators called the central site to receive group assignments.
- The patients in the 2 groups were **similar** with respect to prognostic factors, and this suggested that the randomization was generally successful. There were no statistically significant differences between the 2 groups with respect to demographic characteristics, but there were some statistically significant differences with respect to clinical characteristics. The surgery group had a slightly higher average body mass index than the watchful waiting group (26.6 versus 25.8, $P = 0.004$), fewer sedentary and ambulatory activities, and fewer patients whose hernias had enlarged in the last 6 weeks (9.6% versus 15.4%, $P = 0.04$). How should we interpret these differences between the intervention and control groups? First, it is important to realize that the investigators examined 28 different clinical characteristics. If a P value of 0.05 is used to denote significance, 1 in 20 variables will show statistically significant differences by random chance alone. When many variables are examined, therefore, it is likely that at least 1 variable will show a statistically significant difference. Second, the degree of difference in the variables and the relative influence of a particular prognostic factor on the outcome of interest are more important than the fact that a difference is statistically significant. In an analysis of the issue, it is important as well to determine whether a particular baseline difference would make the study more likely to be positive (favoring the intervention) or negative (favoring the control group). In this study, the small difference in body mass index does not seem likely to have substantially affected the outcome. However, the control group contained 5% more patients whose hernias had enlarged in the last 6 weeks, and this may have had an impact on the outcome. This difference would bias the study against watchful waiting: because the control group, or watchful waiting group, contained more patients whose hernias seemed to be worsening, that group might have been more likely to experience complications. Therefore, if the watchful waiting group did as well

as the surgical group, we might believe even more strongly in the equivalence of the strategies.

- The study was performed and analyzed in an **intention-to-treat fashion**, meaning that all patients were analyzed in the groups to which they were originally assigned. Subjects who were assigned to watchful waiting but who received surgery were counted in the watchful waiting group, and those initially assigned to the surgery group but who did not receive surgery were counted in the surgery group. This approach is important in that it protects the original randomization and maintains the reliability of the results. However, when large numbers of patients cross over, an intention-to-treat analysis may obscure real differences between the 2 treatment strategies because ultimately the groups may have received quite similar treatments. In this study, the rates of crossover between the groups was fairly high (see the Results section), so the investigators performed a secondary analysis based on the treatment that each patient actually received (as-treated or per-protocol analysis). When as-treated analyses are performed, investigators are looking to see if the result is similar to the result of the intention-to-treat analysis to verify that the crossover rate did not largely alter the result.
- The study was not **blinded**. Patients and clinicians obviously could not be blinded to whether surgery was performed. However, the study did blind the outcome assessors, the investigators who determined which patients reached various study endpoints. In a study such as this one with a somewhat subjective primary outcome, the blinding of outcome assessors is an important protection against bias.

Results

Seven hundred twenty-four patients were randomized, 366 to the watchful waiting group and 358 to the surgery group. The mean age of the patients was 57. Two patients from each group were deemed ineligible after randomization and were excluded. Eighty-five patients in the watchful waiting group (23%) received surgery within 2 years, and 62 patients in the surgery group (17%) did not actually undergo surgery.

The primary outcome at 2 years was ascertained in 92.3% of the patients in the watchful waiting group and in 89% of the patients in the surgery group after patients who died, withdrew consent, or were lost to follow-up were excluded. The 2 groups did not differ for the primary outcomes of pain interfering with activities (5.1% in the watchful waiting group versus 2.2% in the surgery group, $P = 0.52$) or a mean physical component score change from the baseline. Complication and death rates were similar in the 2 groups. There were 2 incidents of acute hernia incarceration in the watchful waiting group, representing a rate of a rate of 0.0018 events per patient year. Not surprisingly, patients in the watchful waiting group who ultimately underwent surgery reported increases in pain unpleasantness at the time of crossover. Intention-to-treat and as-treated analyses were similar for the primary outcomes.

Applicability

The surgeon determines that the study in *JAMA* applies to his patient because the patient would have met the entry criteria.

THE PHYSICIAN'S DECISION

The surgeon notes that watchful waiting appears to be a safe strategy in the management of minimally symptomatic inguinal hernias in men. The study by Fitzgibbons and colleagues[2] was well done overall. The surgeon wonders, though, about the impact of the high crossover rate and is particularly surprised by the high number of patients randomized to surgery who did not receive it. He realizes that this may have led to an underestimation of the symptomatic benefits of surgery. However, the surgeon feels comfortable advising his patient that watchful waiting is safe, although the patient may ultimately need surgery.

CONCLUSION

This case illustrates several important points:

- When one is searching for evidence in PubMed/Medline, "Limits" should be used when an initial PubMed/Medline search yields too many articles.
- Randomized trials of surgical interventions can and should be performed. In trials in which patients and their physicians cannot be blinded to group allocation, it is important that outcome assessors be adequately blinded.
- Allocation concealment is an important component of the design of a randomized trial; it minimizes bias that may arise when physicians can predict into which group their patient is likely to be randomized. Allocation concealment is usually accomplished through centralized randomization schemes or computer-generated numbers.

- In a randomized trial, one can test the adequacy of the randomization by looking at whether the groups were similar at the outset of the study with respect to prognostic factors. When many prognostic factors are evaluated, there is likely to be at least 1 factor that shows statistically significant differences. When interpreting these differences, one should focus on the magnitude of the difference, the influence of that particular prognostic factor on the outcome(s) of interest, and the direction in which a real difference would bias the ultimate results of the study.
- When crossover rates are high, investigators can do an as-treated analysis, in which patients are evaluated according to the treatment that they actually received, in addition to the standard intention-to-treat analysis, in which patients are analyzed in the groups to which they were originally randomized. Similar results from the 2 analyses support the notion that the high crossover rate did not substantially impact the result of the study.

In future articles on evidence-based medicine, we will discuss other aspects of the evidence cycle. The complete cycle is shown here:

1. A case and dilemma.
2. A well-formulated clinical question.
3. A search of the literature to answer the question.
4. A critical appraisal of the literature.
5. The incorporation of the patient's values and preferences.
6. Finally, the application of the results to the care of the patient.

Each article in the series will focus on a few specific areas, such as methods for searching the medical literature for meaningful data and the procedure for concealment in medical research. Although not meant to be exhaustive, this series will demonstrate how a physician can obtain reliable, evidence-based answers to specific clinical questions of importance to his or her patients.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 1995; 123: A12–A13.
2. Fitzgibbons RJ Jr, Giobbie-Hurder A, Gibbs JO, et al. Watchful waiting vs repair of inguinal hernia in minimally symptomatic men: a randomized clinical trial. *JAMA* 2006; 295: 285–292.
3. Consort Statement Website. Available at: http://www.consort-statement.org. Accessed July 2008.
4. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1993; 270: 2598–2601.

DOI:10.1002/MSJ

NEW DRUGS AND THE FDA

Hepatitis C: Current and Future Therapies

Poonam Mishra, MD and Donald M. Jensen, MD

Section of Gastroenterology, Hepatology, and Nutrition, Department of Medicine, University of Chicago, Chicago, IL

Chronic hepatitis C is the most common cause of chronic liver disease in the United States. The natural history of hepatitis C involves progression to cirrhosis, hepatocellular carcinoma, liver failure, and death. It is one of the leading indications for liver transplantation and results in approximately 8000 to 10,000 deaths per year.[1] Globally, it is estimated that 175 million persons are infected with hepatitis C virus (HCV), and it affects about 3.2 million people in the United States.[2] On the basis of a computer cohort simulation of the US population between 2010 and 2019, it has been estimated that morbidity and mortality associated with chronic hepatitis C will result in 165,900 deaths from chronic liver disease and 27,200 deaths from hepatocellular carcinoma.[3] Because of its associated morbidity and mortality, it has become imperative that patients be treated effectively with a sustained response rate to minimize the health and economic impacts of this major public health problem.

TREATMENT BACKGROUND

The therapy of chronic hepatitis C has evolved over the last decade, and significant progress has been made since the first introduction of interferon monotherapy in 1986 for the treatment of non-A, non-B hepatitis. Interferon alpha was first improved for the treatment of chronic hepatitis C in 1991. One of the mechanisms involved is up-regulation of interferon-stimulated genes that inhibit HCV replication.[4] Interferon alpha also has an immunomodulatory effect and interacts with both adaptive and innate immune responses of the host.

The addition of ribavirin to interferon was a milestone in the therapy of hepatitis C infection, resulting in better response rates when combination therapy with interferon and ribavirin was used.[5–7] Ribavirin is a nucleoside guanosine analogue, and its antiviral activity involves multiple different potential pathways. It may directly inhibit HCV replication through the inhibition of HCV RNA polymerase. It inhibits the enzyme inosine monophosphate dehydrogenase (IMPDH), a key step in the de novo biosynthesis of guanine nucleotides, and catalyzes the conversion of inosine monophosphate to xanthine monophosphate, leading to a decrease in the intracellular levels of guanosine−5′-triphosphate and 2′-deoxyguanosine 5′-triphosphate. Ribavirin has immunomodulatory effects as well[8] and induces a T helper 1 immune response. Ribavirin may also act by inducing mutations in the viral genome leading to viral mutations that cause replication error catastrophe, which in turn leads to viral extinction.

Therapy is primarily aimed at eradicating HCV and thus preventing disease progression and complications. Antiviral therapy has been shown to favorably alter the natural course of chronic hepatitis C. The long-term virological response is seen in more than 95% of patients who achieve sustained virological response (SVR). SVR has been associated with a decreased progression of fibrosis, and some studies have even suggested reversal of fibrosis or early cirrhosis.[9,10] It has been associated with a reduced incidence of hepatocellular carcinoma, particularly among patients who achieve SVR.[11] It

Address Correspondence to:

Donald M. Jensen
Center for Liver Diseases
University of Chicago
Chicago, IL
Email: djensen@medicine.bsd.uchicago.edu

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20060

© 2008 Mount Sinai School of Medicine

has also shown a survival benefit. However, the overall response rate is only about 40% to 45% for genotype 1 patients and 75% to 80% for genotype 2 and 3 patients when they are treated with pegylated interferon in combination with ribavirin.[12–14]

TRENDS TO INDIVIDUALIZED TREATMENT

Currently, the US Food and Drug Administration–approved standard treatment of chronic hepatitis C is combination therapy with either peginterferon Alfa-2a or Alfa-2b plus ribavirin for a total duration of 24 or 48 weeks according to the genotype.[15]

In this evolving era of individualized therapy, emphasis is now placed on tailoring therapy on the basis of host and viral factors rather than a one-size-fits-all approach. Factors that have been shown to predict a lower SVR rate include genotype 1, a high viral load at baseline, advanced disease on histology (bridging fibrosis and cirrhosis), obesity, older age, and African American race.[16]

Recent studies have aimed to individualize therapy on the basis of the virological response at week 4 of therapy. Rapid virological response (RVR) is defined as undetectable HCV RNA at week 4 of therapy and is likely more accurate than baseline factors in predicting SVR.[17] In a recent study, Jensen and colleagues[18] showed that patients with genotype 1 who have an RVR can achieve an SVR with just 24 weeks of therapy equal to that achieved with 48 weeks. Conversely, patients with genotype 1 who are slow responders (HCV RNA detectable at week 4) may benefit from an extension of the total duration of therapy to 72 weeks, but this needs to be confirmed.[19] Weight-based ribavirin may be beneficial for genotype 1 patients (1000 mg for less than 75 kg, 1200 mg for 75–104 kg, and up to 1400 mg/day for 105–125 kg).[20] A recently completed clinical trial evaluating the use of high-dose peginterferon, with or without high-dose ribavirin, in genotype 1 patients with a high viral load who are greater than 85 kg in weight has been reported.[21] This study has demonstrated the advantage of both higher dose (270 μg) peginterferon and 1600 mg of ribavirin in producing an SVR in this difficult-to-treat group.

In a large, randomized, multinational trial by Shiffman and colleagues[22] in patients infected with HCV genotype 2 or 3, treatment with peginterferon and ribavirin for 16 weeks resulted in a lower overall SVR rate and higher relapse rates in comparison with those of patients treated with the standard 24-week regimen. In 2 previous studies performed in patients with genotypes 2 and 3 with a low viral load who achieved RVR (undetectable HCV RNA at week 4), 12 to 16 weeks of therapy with pegylated interferon and ribavirin was reported to be sufficient.[23,24] In genotype 3 patients with a high viral load (baseline viral load above 800,000 IU/mL), a longer duration of treatment, that is, 24 weeks, resulted in improved SVR rates in comparison with 16 weeks of therapy, although this did not reach statistical significance.[24]

In a recent study,[25] investigators reported that in patients with chronic hepatitis C genotype 4 who achieve undetectable HCV RNA at week 4, 24 weeks of therapy is sufficient. However, in patients who have undetectable HCV RNA at week 12, extending therapy to a total duration of 36 weeks is beneficial.

The use of growth factors to support the therapy and to minimize dose reductions may potentially lead to better response rates, but data from large randomized clinical trials showing the efficacy of growth factors are still pending. Furthermore, the use of erythropoietin is now under the shadow of a black box warning for the risks of cardiovascular complications and potential for red cell aplasia. A recent study reported that using epoetin alpha in all subjects at the initiation of therapy with pegylated interferon and ribavirin did not enhance SVR, but subjects who were given higher doses of ribavirin had increased SVR because of significantly lower relapse rates, and such patients may need epoetin alpha at the initiation of therapy.[26] A previous study by Afdhal *et al.*[27] reported that the use of epoetin alpha helped to maintain the ribavirin doses and improved the quality of life and hemoglobin levels versus those of a placebo group. Data from the lead-in phase of the Hepatitis C Long-Term Treatment Against Cirrhosis Trial reported that reducing the dose of peginterferon had a minimal effect on SVR, whereas reducing the dose of ribavirin during the first 20 weeks of therapy adversely affected SVR.[28] However, a recent study of patients who were previous nonresponders to standard interferon with or without ribavirin when retreated with peginterferon alpha-2a and ribavirin demonstrated that reducing the cumulative dose of peginterferon during the first 20 weeks of therapy adversely affected the viral clearance and SVR. When the patients remained on full-dose peginterferon during this period, a reduction in ribavirin did not affect viral clearance and SVR. However, premature discontinuation of ribavirin therapy was found to be associated with decreased viral clearance and SVR.[29] These studies support the notion that ribavirin dosing and adherence are important throughout the treatment period and that every attempt should be made to avoid significant and prolonged dose reductions and, in particular, interruptions.

There are some recent data suggesting that a subset of patients with cirrhosis and decompensated disease may benefit from antiviral therapy. Response rates in the range of 20% to 30% (13% in genotype 1 patients and up to 50% in non-1 genotype patients) have been reported. In patients with cirrhosis awaiting liver transplantation, a low accelerating dosage regimen with peginterferon and ribavirin has been used.[30] Investigators in this study concluded that this may help to stabilize the disease course and may prevent the recurrence of HCV after transplantation.

TREATMENT OF EXPERIENCED PATIENTS

In recent years, the number of patients who have been nonresponders to previous therapy with either interferon with ribavirin or peginterferon with ribavirin has increased, and it is has been speculated that it will soon surpass the number of treatment-naïve patients. There is an unmet need to treat this subset of patients, and it continues to remain a challenging endeavor.

Patients who have relapsed after previous therapy with interferon and ribavirin have better response rates than true nonresponders when they are retreated with peginterferon and ribavirin; however, the number of studies in this patient population is small. SVR rates are better in relapsers previously treated with standard interferon and ribavirin therapy and retreated with pegylated interferon plus ribavirin: they range from 32% to 50%.[31] A multinational study of patients who previously relapsed during a 24-week follow-up period after achieving an end-of-treatment response at 24 weeks with peginterferon alpha-2a plus ribavirin combination therapy reported an overall SVR rate of 55% when the patients were retreated with 48 weeks of peginterferon alpha-2a plus ribavirin.[32] These results hold promise for the subset of patients who relapse after 24 weeks of therapy or who have to discontinue therapy prematurely because of side effects.

Studies have demonstrated that patients who are prior nonresponders to interferon and ribavirin therapy, when retreated with pegylated interferon plus ribavirin, achieve SVR rates in the range of 6% to 18%.[28] Treatment failure in these patients may result from several factors that influence response to treatment, including an inadequate therapeutic regimen, viral factors, and host factors. Some of these are nonmodifiable factors, including race, age, viral genotype, baseline viral load, sensitivity to interferon, disease factors such as cirrhosis, transplantation, and coinfection with human immunodeficiency virus or chronic hepatitis B virus.[15,33–39] Factors that are modifiable include adherence and compliance to therapy, length of treatment duration based on genotype, interferon and ribavirin dosage, body mass index/body weight, and insulin resistance.[15,33,35–40] Identifying the reason for primary nonresponse and relapse in these patients is the key to successful retreatment.

NOVEL TREATMENTS

Significant advancements have been made in the development of newer drugs for the treatment of HCV, but at this time, it appears that interferon will still remain the mainstay of therapy for HCV in the future in combination with other drugs. Emphasis is now on developing forms of interferon that are better tolerated, have longer duration of action, and have increased efficacy.

Albuferon (Albinterferon Alfa-2b)

Albinterferon is a fused form of interferon alpha 2b with human serum albumin. Data so far suggest that it is well tolerated and has a longer half-life than pegylated interferon. Interim results from phase 2b trials have suggested that albinterferon has antiviral activity and safety comparable to those of pegylated interferon with a better dosing schedule.[41] However, higher rates of mild cough and dyspnea were reported in the albinterferon group, and discontinuations were higher in the 1200-μg treatment group. Large, multicenter phase III studies are underway in genotype 1 and genotype 2 or 3 patients with a dose reduction in the Albuferon dose from 1200 to 900 μg because of concerns related to pulmonary complications.[42]

Consensus Interferon

In an open-label pilot study of hepatitis C patients who were previous nonresponders to interferon and ribavirin therapy, investigators reported an SVR of 22% when daily consensus interferon was used in combination with 1000/1200 mg of ribavirin. Authors reported that an 8-week induction regimen of higher doses of consensus interferon did not increase the SVR rates in comparison with 9 μg of consensus interferon for 48 weeks and hence suggested that an induction therapy is not required.[43] This may be a potential new therapeutic option for the hard-to-treat nonresponder patient population, but data presented by Bacon *et al.*[44] revealed only a 5% to 10% SVR rate in prior peginterferon and ribavirin null responders, a particularly difficult patient population.[44]

RIBAVIRIN ANALOGUES AND INOSINE MONOPHOSPHATE DEHYDROGENASE INHIBITORS

A major limitation associated with the use of ribavirin is hemolytic anemia. This may result in dose reductions or discontinuation of therapy and thus lower response rates. Recently, there has been work on developing newer ribavirin analogues and IMPDH inhibitors.

Taribavirin (Viramidine, Valeant Pharmaceuticals International, Aliso Viejo, CA)

Viramidine is a prodrug of ribavirin. Viramidine failed to show any improved efficacy over ribavirin in a phase 2 study, although the side effect of hemolytic anemia was reported to be low. In an intent-to-treat analysis done for a phase 3 trial, viramidine did not meet the noninferiority to ribavirin efficacy.[45] Nonetheless, weight-based dosing of viramidine is currently being tested in clinical trials to determine the most appropriate dosing strategy.

Merimepodib (VX-497)

Merimepodib, an inhibitor of IMPDH, was studied in combination with pegylated interferon and ribavirin. Preliminary data from the phase 2b trial demonstrating a lack of efficacy resulted in the discontinuation of further clinical trials.

Mizoribine is an IMPDH inhibitor that has been shown to inhibit viral replication, but it is not being studied under clinical trials for HCV therapy. Mycophenolate mofetil, also an IMPDH inhibitor, is currently used as an immunosuppressive agent and is currently not being studied for its antiviral property.[46]

SMALL-MOLECULE INHIBITORS

The concept of specifically targeted antiviral therapy for HCV is being discussed widely, with the main focus on increasing SVR with a shorter duration of therapy so that side effects can be minimized. Recent advances in the understanding of HCV replication and virally encoded enzymes have opened the path toward candidate drugs targeting nonstructural protein 3/4A protease and nonstructural protein 5B RNA–dependent RNA polymerases. The most promising advances for novel therapies for the future include protease and polymerase inhibitors that inhibit HCV replication (Table 1).

Hepatitis C Virus Protease Inhibitors

Many products are in the pipeline of drug discovery and are in the early phases of development. At least one is currently entering a phase III clinical trial.

BILN 2061 (Boehringer Ingelheim)

This drug reduced HCV RNA levels in 2 days of treatment and was reported to be well tolerated in a small study, but further clinical trials in humans were halted because of cardiac toxicity seen in monkeys when higher doses of the drug were given for a longer duration.[47]

Telaprevir (VX-950, Vertex Pharmaceuticals, Cambridge, MA)

Promising results have been reported for serine protease inhibitor VX-950. Initial data reported from a phase 1b, placebo-controlled, double-blind randomized study using monotheraphy for 14 days reported median reduction of 4.4 log 10 in HCV RNA when 750 mg of the drug was administered orally every 8 hours."[48] Results from two phase 2 studies in treatment-naïve genotype 1 HCV patients were recently reported. Patients in PROVE-1: A Phase 2 Study of Telaprevir with Peginterferon alfa-2a and Ribavirin in Treatment Naive Subjects with Hepatitis C" were assigned to 4 different treatment arms, 3 of which received telaprevir and 1 of which received the standard therapy for 48 weeks. All 3 telaprevir arms received 750 mg of telaprevir every 8 hours for 12 weeks in addition to pegylated interferon Alfa-2a (180 μg/week) and ribavirin (1000–12,000 mg/day). One arm stopped all 3 drugs at 12 weeks, the second arm continued pegylated interferon and ribavirin for an additional 12 weeks, and the third arm continued pegylated interferon and ribavirin for an additional 36 weeks. Patients in all 3 arms who received telaprevir had a significantly higher rate of RVR or undetectable HCV RNA at week 4 in comparison with the control group, and the same was seen in terms of early virological response at week 12. Investigators have suggested that the addition of telaprevir may result in improved SVR rates with a shorter duration of therapy. However, patients who received less than 24 weeks of therapy (ie, 12 weeks) had an increased relapse rate. Additionally, more

Table 1. *Specifically Targeted Antiviral Therapy for HCV: Drugs in Development.*

Drug Class	Drug	Mean Log Reduction in HCV RNA	Patients Being Studied	Current Phase of the Trial	References
Enzyme Inhibitors					
Protease inhibitors	Telaprevir (VX-950)	A 2–4.4 log 10 reduction in 14 days when it was used as monotherapy in N patients. Improved SVR rates with triple therapy.	N/TE	Phase III	Reesink *et al.*[48] McHutchison *et al.*[49] Dusheiko *et al.*[50]
	Boceprevir (SCH 503034)	1.0–1.6 log IU/mL in 7 days when it was used as monotherapy in previous nonresponders. Increased RVR and EVR in the lead-in-arm.	N/TE	Phase II	Zeuzem *et al.*[51] Kwo *et al.*[52] Schiff *et al.*[53]
	ITMN-191	Dose-dependent 1.6–3.9 log reduction in 14 days.	N	Phase I	InterMune[55]
Polymerase inhibitors	R1626 (Roche)	Dose-dependent 1.2–3.7 log reductions when it was given for 14 days. Eighty-four percent EOT response with P & R versus 65% in the control group.		Phase II	Roberts *et al.*[56] Nelson *et al.*[57]
	R7128 (Pharmasset)	A 2.7 log 10 drop with a 1500-mg twice daily dose for 14 days when it was used as monotherapy in patients who failed with interferon previously. Eighty-five percent achieved RVR in 4 weeks when it was used in combination with P and R.	N/TE	Phase I	Reddy *et al.*[58] Pharmasset[59]
Interferon and R Modifications					
IMPDH inhibitors	Albuferon (albinterferon Alfa-2b)	Antiviral activity comparable to pegylated interferon with a better dosing schedule.	N/TE	Phase III	Human Genome Sciences[41,42]
	Infergen (consensus interferon)	An SVR rate of 5–10% in prior P and R null responders.	TE	Phase IV	Cornberg *et al.*[43] Bacon *et al.*[44]
	Taribavirin (viramidine)	It did not meet the noninferiority efficacy goal. Lower rate of hemolytic anemia.	N	Phase IIb	Valeant[45]
Targeting Cellular factors					
Cyclophilin inhibitor	Debio 025	Dose-dependent 4.6–4.8 log 10 reduction in combination with P. A 2.20 log 10 reduction in the Debio 1000 mg/day monotherapy arm.	N	Phase II	Flisiak *et al.*[64]
Thiazolides	Nitazoxanide	SVR rates of 79% and 25% in the N group and TE group genotype 4 patients, respectively	N/TE	Phase II	Rossignol *et al.*[62]
Caspase inhibitors	IDN-6556	Data are limited. A serum enzyme level reduction has been reported.		Phase II	Pockros *et al.*[66]

adverse events were noted in the telaprevir arms in comparison with standard therapy, leading to more treatment discontinuations.[49] Interim results from another study (PROVE-2: Treatment of Chronic Hepatitis C with Telaprevir (TVR) in Combination with Peginterferon Alfa-2a with or without Ribavirin) reported lower HCV RNA suppression and higher relapse rates in the arm receiving telaprevir and pegylated interferon without ribavirin.[50] The US Food and Drug Administration has approved the fast-track designation for VX-950.

Boceprevir (SCH 503 034, Schering-Plough, Kenilworth, NJ)

Boceprevir is an oral HCV NS3 serine protease inhibitor, and it has been studied as monotheraphy in genotype 1 patients who are nonrespondres to previous therapy with pegylated interferon. Dose-dependent reductions in HCV RNA levels have been reported in a 14-day study.[51] Phase 2 trials are being conducted with boceprevir monotherapy as well as boceprevir in combination with pegylated interferon. Boceprevir plus peginterferon combination therapy has shown potent antiviral activity. This drug is being proposed for patients with HCV genotype 1 who have been nonresponders to previous therapy with pegylated interferon and ribavirin and has been granted the fast-track designation.

A novel approach was used in the phase 2 study performed in treatment-naïve genotype 1 HCV patients, which assessed the use of 800 mg of boceprevir twice daily with peginterferon and ribavirin. One of the arms had a lead-in phase in which patients received 4 weeks of pegylated interferon Alfa-2b plus ribavirin (800–1400 mg/day) followed by triple therapy for 24 to 44 weeks. The rationale behind this approach is the addition of a protease inhibitor after steady-state levels of pegylated interferon and ribavirin have been reached and the immune system has been primed by interferon. This approach, it is hoped, may reduce the risk of drug resistance while potentiating antiviral activity. Increased RVR and early virological response as well as reduced viral breakthrough were reported in the lead-in arm in the interim analysis presented from this ongoing study.[52] Additionally, data from a study evaluating boceprevir in previous null-responder patients also look promising.[53]

The emergence of resistant mutations when these compounds are used as monotherapy has been a significant issue. In vitro resistance mutations against VX-950 and BILN 2061 have already been described. Recently, mutations that are resistant to both HCV protease inhibitors have been identified.[54]

ITMN-B (Intermune)

ITMN-191 is a HCV protease inhibitor that is being studied both in combination with polymerase inhibitor R1479 and in combination with interferon.

In a phase 1b trial using ITMN-191 as monotherapy in chronic hepatitis C genotype 1 patients, a dose-related response reduction of HCV RNA was reported.[55] Further research trials evaluating 14-day triple combination therapy are planned in the future.

Hepatitis C Virus Polymerase Inhibitors

R1626 (Roche)

R1626, an oral prodrug of a nucleoside analogue targeting the HCV polymerase enzyme, has shown potent antiviral activity and is well tolerated in naïve patients. It is the most advanced drug under development in this class, and the reported viral activity is comparable to that of the protease inhibitors being studied. No evidence of viral resistance has yet been demonstrated. Dose-dependent antiviral activity was demonstrated when up to 4500 mg of R1626 was given for 14 days.[56] Anemia and cytopenia were reported with the use of the drug and were reversible. Data from the phase 2a study using 4 weeks of R1626 in combination with pegylated interferon Alfa-2a versus triple therapy with R1626 (1500 mg twice daily) plus pegylated interferon Alfa-2a and ribavirin were recently reported. After 4 weeks, all patients received pegylated interferon and ribavirin for 44 additional weeks. HCV RNA decreased by a mean of 5.2 log 10 from the baseline in the triple therapy.[57]

R7128 (Pharmasset, Inc.)

R7128 is a nucleoside analogue polymerase inhibitor and is currently being studied in combination with peginterferon Alfa-2a and ribavirin in both treatment-naïve and treatment-experienced patients. In a phase 1 trial of HCV genotype patients who previously failed interferon therapy, dose-dependent antiviral activity was reported with a mean 2.7 log 10 decrease in HCV RNA when R7128 (1500 mg twice daily) was used for 14 days as monotherapy.[58] Four-week data from an ongoing phase 1 study in naïve genotype 1 patients using R7128 in combination with pegylated interferon alpha-2a and ribavirin reported an 85% RVR.[59]

Valopicitabine (NM 283, Idenix)

Valopicitabine is an investigational NS5B polymerase inhibitor that was studied as monotherapy and in combination with pegylated interferon and ribavirin. Despite earlier results demonstrating additive antiviral activity, the incidence of gastrointestinal-related adverse events and laboratory abnormalities (sporadic elevations of aspartate aminotransferase, alanine aminotransferase, pancreatic amylase and lipase) was higher in the treatment arm. Hence, the development of this molecule was discontinued because of concerns over the overall risk/benefit ratio.

HCV-796 (ViroPharma and Wyeth)

HCV-796 is a nonnucleoside polymerase inhibitor. Data from the phase 1b study using HCV-796 in the dose range of 100 to 1000 mg twice daily in combination with pegylated interferon Alfa-2b over a 14-day period reported a mean HCV RNA reduction of 3.3 to 3.5 log10.[60] The phase 2 study using HCV-796 in combination with pegylated interferon and ribavirin was discontinued because of elevated liver enzymes noted in some patients after 8 weeks or more of therapy.

NITAZOXANIDE

Nitazoxanide is a drug from the class of thiazolides whose mechanism of action is unknown, but it is thought to activate an innate defense against viruses. It has fewer side effects than interferon, with an added advantage of being an oral drug. It was evaluated in chronic hepatitis C patients with genotype 4. Nitazoxanide was given as monotherapy (500 mg twice daily) for 12 weeks; it was followed by nitazoxanide plus pegylated interferon for 36 weeks in one arm, and the other arm received nitazoxanide with pegylated interferon and ribavirin for 36 weeks. The third arm was the standard of care. Investigators reported improved SVR rates in arms receiving nitazoxanide versus the standard of care, with the triple therapy arm doing better than the dual therapy arm.[61,62] Further data reported from a study using 4-week lead-in nitazoxanide monotherapy demonstrated comparable results, with the lead-in phase being shortened from 12 to 4 weeks.[63] Further trials are planned in the United States to evaluate the efficacy and safety of nitazoxanide in treatment-naïve patients with genotype 1.

CYCLOPHILIN INHIBITOR DEBIO 025

The safety and efficacy of Debio 025, a selective cyclophilin inhibitor, was evaluated in a phase II study in combination with pegylated interferon in chronic hepatitis C patients.

Debio 025 at doses of 600 and 1000 mg daily for 29 days had an additive anti-HCV effect when coadministered with pegylated interferon alpha-2a in treatment-naïve HCV patients.[64]

TOLL-LIKE RECEPTOR AGONISTS

Recently, data from a phase 1B, randomized, double-blind, dose-escalation trial of CPG 10101, a toll-like receptor 9 agonist, in 60 patients with HCV were reported. Dose-dependent cytokine induction was seen, resulting in increased markers of immune activation and decreases in HCV RNA levels.[65]

CASPASE INHIBITORS

Caspase inhibitors are being evaluated as adjuncts to delay the development of apoptosis, which is a major player in liver injury. Dada is very limited, and a reduction in serum aminotransferase liver enzymes was reported when IDN-6556 was used in HCV patients.[66]

THERAPEUTIC VACCINES

Several efforts have been made in the development of a therapeutic vaccine, but the results so far have not been encouraging. The most recent development in the field is the novel approach of a dendritic cell vaccine. Chronic hepatitis C infection results in inefficient antiviral T cell responses. Dendritic cells are potent antigen-presenting cells and hence have the ability to prime and activate T cells. This has been targeted as a potential approach to the development of therapeutic and prophylactic vaccines.

Investigators are using different approaches to develop this therapeutic vaccine. Dendritic cells are pulsed with HCV core proteins and peptides or are infected with adenoviral vectors containing nonstructural proteins (NS3, NS4 and NS5B) or HCV core genes. Strong immune responses have been reported with these vaccines. An in vivo study in mice showed partial protection.[67] More recently, an HCV T cell–based, modified virus Ankara–vectored vaccine was studied in a transgenic mouse model and was reported to induce potent long-lasting and in vivo cross-reactive T cell responses. It is being studied in clinical trials now. Further studies are warranted to explore the field of HCV therapy with immunomodulation.

SHORT-INTERFERING RNA (SiRNA) HEPATITIS C VIRUS (HCV) INTERNAL RIBOSOME ENTRY SITE (IRES) INHIBITORS

ISIS 14803 is an antisense oligonucleotide that has been shown to inhibit protein expression in HCV cell culture and mouse models. It binds to the HCV RNA at the translation initiation region of the IRES. An initial open-label, dose-escalation study of this agent failed to show any reduction in HCV RNA levels over 4 weeks.[68] Further studies are being conducted to evaluate the safety and efficacy of this agent. Short-interfering RNAs are currently in development.[69]

CONCLUSION

There is clearly an unmet need for new therapeutic modalities for the treatment of chronic HCV infection that are more efficacious with a shorter duration of therapy and improved tolerability. Numerous small-molecule agents targeting different critical phases in the viral life cycle are under development and are being currently evaluated in clinical trials. At this point, it seems that interferon will still remain the mainstay of treatment in combination with small molecules from different classes to prevent the emergence of drug-resistant variant strains, although promising data from the PROVE-1 rollover: A Study of Telaprevir (TVR) with Peginterferon Alfa-2a (P) and Ribavirin (R) in Subjects with Well-documented Prior P/R Null Response, Non-Reponse or Relapse. in peginterferon/ribavirin nonresponders[70] might question this assumption. Despite initial encouraging data, many questions still remain unanswered, and we will have to wait for the results of large randomized placebo-controlled clinical trials before the use of any of these agents in clinical practice can be advocated. The future holds the promise and hope that we will be able to tailor our therapies on the basis of host and viral factors with improved clinical outcomes, hence lowering the huge economic burden imposed by this chronic disease.

DISCLOSURES

Potential conflict of interest: Nothing to report.

DOI:10.1002/MSJ

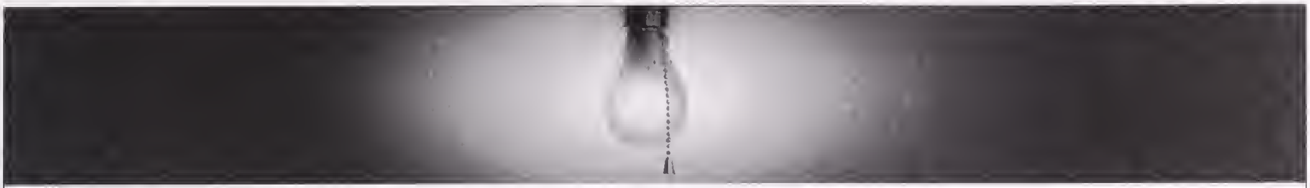
REFERENCES

1. National Institutes of Health consensus development conference statement: management of hepatitis C: 2002. *Gastroenterology* 2002; 123: 2082–2099.
2. Armstrong GL, Wasley A, Simard EP, et al. The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. *Ann Intern Med* 2006; 144: 705–714.
3. American Gastroenterological Association. Technical review on the management of hepatitis C. *Gastroenterology* 2006; 130: 231–264.
4. Hoofnagle JH, Seefe LB. Peginterferon and ribavirin for chronic hepatitis C. *N Engl J Med* 2006; 355: 2444–2451.
5. Schalm SW, Hansen BE, Chemello L, et al. Ribavirin enhances the efficacy but not the adverse effects of interferon in chronic hepatitis C. Meta-analysis of individual patient data from European centers. *J Hepatol* 1997; 26: 961–966.
6. Schvarcz R, Yun ZB, Sonnerborg A, Weiland O. Combined treatment with interferon alpha-2b and ribavirin for chronic hepatitis C in patients with a previous non-response or non-sustained response to interferon alone. *J Med Virol* 1995; 46: 43–47.
7. Reichard O, Norkrans G, Fryden A, et al., and the Swedish Study Group. Randomized, double-blind, placebo-controlled trial of interferon &agr -2b with and without ribavirin for chronic hepatitis C. *Lancet* 1998; 351: 83–87.
8. Rigopoulou EI, Abott WG, Williams R, Naoumov NV. Direct evidence for immunomodulatory properties of ribavirin on T-cell reactivity to hepatitis C virus. *Antiviral Res* 2007; 75: 36–42.
9. Poynard T, McHutchison J, Davis GL, et al. Impact of interferon Alfa-2b and ribavirin on progression of liver fibrosis in patients with chronic hepatitis C. *Hepatology* 2000; 32: 1131–1137.
10. Poynard T, McHutchison J, Manns M, et al. Impact of pegylated interferon Alfa-2b and ribavirin on liver fibrosis in patients with chronic hepatitis C. *Gastroenterology* 2002; 122: 1303–1313.
11. Shiratori Y, Ito Y, Yokosuka O, et al. Antiviral therapy for cirrhotic hepatitis C: association with reduced hepatocellular carcinoma development and improved survival. *Ann Intern Med* 2005; 142: 105–114.
12. Manns MP, McHutchison JG, Gordon SC, et al. PEG-interferon Alfa-2b in combination with ribavirin compared to interferon Alfa-2b plus ribavirin for initial treatment of chronic hepatitis C. *Lancet* 2001; 358: 958–965.
13. Fried MW, Shiffman ML, Reddy KR, et al. Peginterferon Alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* 2002; 347: 975–982.
14. Hadziyannis SJ, Sette H Jr, Morgan TR, et al., and the PEGASYS International Study Group. Peginterferon-&agr 2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. *Ann Intern Med* 2004; 140: 346–355.
15. Strader DB, Wright T, Thomas DL, Seeff LB, and the American Association for the Study of Liver Diseases. Diagnosis, management and treatment of hepatitis C. *Hepatology* 2004; 39: 1147–1171.
16. Zeuzem S. Heterogenous virologic response rates to interferon-based therapy in patients with chronic

hepatitis C: who responds less well? *Ann Intern Med* 2004; 140: 370–381.

17. Shiffman ML, Chung RT, Hamzeh FM. Time to HCV RNA undetectability supersedes baseline factors in predicting SVR in patients with HCV genotype 1. *J Hepatol* 2008; 48(suppl 2): S313.
18. Jensen DM, Morgan TR, Marcellin P, et al. Early identification of HCV genotype 1 patients responding to 24 weeks peginterferon &agr -2a (40kd)/ribavirin therapy. *Hepatology* 2006; 43: 954–960.
19. Sanchez-Tapias JM, Diago M, Escartin P, et al. Peginterferon-Alfa2a plus ribavirin for 48 versus 72 weeks in patients with detectable hepatitis C virus RNA at week 4 of treatment. *Gastroenterology* 2006; 131: 451–460.
20. Jacobson IM, Brown RS Jr, Freilich B, et al. Peginterferon Alfa-2b and weight-based or flat-dose ribavirin in chronic hepatitis C patients: a randomized trial. *Hepatology* 2007; 46: 971–981.
21. Freid M, Jensen D, Rodriguez-Torres M, et al. Improved sustained virological response (SVR) rates with higher, fixed doses of peginterferon Alfa-2a (40KD) (PEGASYS®) plus ribavirin (RBV) (COPEGUS®) in patients with "difficult-to-cure" characteristics [abstract]. *Hepatology* 2006; 44(suppl 1): 314A.
22. Shiffman ML, Suter F, Bacon BR, et al. Peginterferon Alfa-2a and ribavirin for 16 or 24 weeks in HCV genotype 2 or 3. *N Engl J Med* 2007; 357: 124–134.
23. Mangia A, Santoro R, Minerva N, et al. Peginterferon Alfa-2b and ribavirin for 12 vs. 24 weeks in HCV genotype 2 or 3. *N Engl J Med* 2005; 352: 2609–2617.
24. von Wagner M, Huber M, Berg T, et al. Peginterferon-alpha-2a (40KD) and ribavirin for 16 or 24 weeks in patients with genotype 2 or 3 chronic hepatitis C. *Gastroenterology* 2005; 129: 522–527.
25. Kamal SM, El Kamary SS, Shardell MD, et al. Pegylated interferon alpha-2b plus ribavirin in patients with genotype 4 chronic hepatitis C: the role of rapid and early virologic response. *Hepatology* 2007; 46(suppl 1): 1732–1740.
26. Shiffman ML, Salvatore J, Hubbard S, et al. Treatment of chronic hepatitis C virus genotype 1 with peginterferon, ribavirin, and epoetin alpha. *Hepatology* 2007; 46: 371–379.
27. Afdhal NH, Dieterich DT, Pockros PJ, et al., and the Proactive Study Group. Epoetin Alfa maintains ribavirin dose in HCV-infected patients: a prospective, double-blind, randomized controlled study. *Gastroenterology* 2004; 126: 1302–1311.
28. Shiffman ML, Di Bisceglie AM, Lindsay KL, et al. Peginterferon Alfa-2a and ribavirin in patients with chronic hepatitis C who have failed prior treatment. *Gastroenterology* 2004; 126: 1015–1023.
29. Shiffman ML, Ghany MG, Morgan TR, et al. Impact of reducing peginterferon Alfa-2a and ribavirin dose during retreatment in patients with chronic hepatitis C. *Gastroenterology* 2007; 132: 103–112.
30. Everson GT, Trotter J, Forman L, et al. Treatment of advanced hepatitis C with a low accelerating dosage regimen of antiviral therapy. *Hepatology* 2005; 42: 255–262.
31. Keeffe EB. Chronic hepatitis C: management of treatment failures. *Clin Gastroenterol Hepatol* 2005; 3: S102–S105.
32. Berg C, Goncales FL Jr, Bernstein DE, et al. Retreatment of chronic hepatitis C patients after relapse: efficacy of peginterferon-alpha-2a(40kDa) and ribavirin. *J Viral Hepat* 2006; 13: 435–440.
33. Dienstag JL, McHutchison JG. American Gastroenterological Association medical position statement on the management of hepatitis C. *Gastroenterology* 2006; 130: 225–230.
34. Conjeevaram HS, Freid MW, Jeffers LJ, et al. Peginterferon and ribavirin treatment in African American and Caucasian American patients with hepatitis c genotype 1. *Gastroenterology* 2006; 131: 470–477.
35. Pawlotsky JM. Current and future concepts in hepatitis C therapy. *Semin Liver Dis* 2005; 25: 72–83.
36. Patel K, McHutchison JG. Current therapies for chronic hepatitis C. *Postgrad Med* 2003; 114: 48–52, 57–59, 62.
37. McHutchison JG, Manns M, Patel K, et al. Adherence to combination therapy enhances sustained response in genotype-1-infected patients with chronic hepatitis C. *Gastroenterology* 2002; 123: 1061–1069.
38. Hoefs J, Aulakh VS. Treatment of chronic HCV infection in special populations. *Int J Med Sci* 2006; 3: 69–74.
39. Bacon BR, McHutchison JG. Into the light: strategies for battling hepatitis C. *Am J Manag Care* 2007; 13: S319–S326.
40. Romero-Gomez M, Del Mar Viloria M, Andrade RJ, et al. Insulin resistance impairs sustained response rate to peginterferon plus ribavirin in chronic hepatitis C patients. *Gastroenterology* 2005; 128: 636–641.
41. Human Genome Sciences. Press release (November 5, 2007). Available at: http://www.hgsi.com. Accessed July 2008.
42. Human Genome Sciences. Human genome sciences modifies dosing in ACHIEVE trials of Albuferon (press release, January 23, 2008). Available at: http://www.hgsi.com. Accessed July 2008.
43. Cornberg M, Hadem J, Hermann E, et al. Treatment with daily consensus interferon (CIFN) plus ribavirin in non-responder patients with chronic hepatitis C: a randomized open-label pilot study. *J Hepatol* 2006; 44: 291–301.
44. Bacon B, Regev A, Ghalib RH, et al. The DIRECT trial (daily-dose consensus interferon and ribavirin: efficacy of combined therapy): treatment of non-responders to previous pegylated interferon plus ribavirin: sustained virologic response data [abstract]. *Hepatology* 2007; 46(suppl 1): 311A.
45. Valeant Pharmaceuticals. Valeant Pharmaceuticals reports VISER 2 results for viramidine (press release, September 12, 2006). Available at: http://www.valeant.com. Accessed July 2008.
46. Gish RG. Treating HCV with ribavirin analogues and ribavirin-like molecules. *J Antimicrob Chemother* 2006; 57: 8–13.
47. Hinrichsen H, Benhamou Y, Wedmeyer H, et al. Short-term antiviral efficacy of BILN 2061, a hepatitis C virus serine protease inhibitor, in hepatitis C genotype 1 patients. *Gastroenterology* 2004; 127: 1347–1355.
48. Reesink HW, Zeuzem S, Weegink CJ, et al. Final results of a phase 1B, multiple-dose study of VX-950, a hepatitis C virus protease inhibitor [abstract]. *Hepatology* 2005; 42: 234A–235A.
49. McHutchison JG, Everson GT, Gordan SC, et al. PROVE 1: results from a phase 2 study of telaprevir with peginterferon Alfa-2a and ribavirin in treatment-naïve subjects with hepatitis C. *J Hepatol* 2008; 48(suppl 2): S4.

50. Dusheiko GM, Hezode C, Pol S, et al. Treatment of chronic hepatitis C with telaprevir (TVR) in combination with peginterferon-Alfa-2a with or without ribavirin: further interim analysis of the PROVE 2 study. *J Hepatol* 2008; 48(suppl 2): S26.
51. Zeuzem S, Sarrazin C, Rouzier R, et al. Anti-viral activity of SCH 503034, a HCV protease inhibitor, administered as monotherapy in hepatitis C genotype-1 patients refractory to pegylated interferon (PEG-IFN-Alfa) [abstract]. *Hepatology* 2005; 42(suppl): 233A–234A.
52. Kwo P, Lawitz E, McCone J, et al. Interim results from HCV SPRINT-1: RVR/EVR from phase 2 study of boceprevir plus PegINTRON (peginterferon Alfa-2b)/ribavirin in treatment-naïve subjects with genotype-1 CHC. *J Hepatol* 2008; 48(suppl 2): S372.
53. Schiff E, Poordad F, Jacobson I, et al. Boceprevir (B) combination therapy in null responders (NR): response dependent on interferon responsiveness. *J Hepatol* 2008; 48(suppl 2): S46.
54. Lin C, Gates CA, Rao BG, et al. In vitro studies of cross-resistance mutations against two hepatitis C virus serine protease inhibitors, VX-950 and BILN 2061. *J Biol Chem* 2005; 280: 36784–36791.
55. InterMune, Inc. InterMune announces top-line results of phase 1b trial of ITMN-191 (R7227) and provides program update (press release, April 1, 2008). Available at: http://www.intermune.com. Accessed July 2008.
56. Roberts S, Cooksley G, Shaw D, et al. Interim results of a multiple dose ascending study of R1626, a novel nucleoside analog targeting HCV polymerase in chronic HCV patients. *J Hepatol* 2006; 44(suppl 2): S269.
57. Nelson D, Pockros PJ, Godofsky E, et al. High end-of-treatment response (84%) after 4 weeks of R1626, peginterferon Alfa-2a (40kd) and ribavirin followed by a further 44 weeks of peginterferon Alfa-2a and ribavirin. *J Hepatol* 2008; 48(suppl 2): S371.
58. Reddy R, Rodriguez-Torres M, Gane E, et al. Antiviral activity, pharmacokinetics, safety, and tolerability of R7128, a novel nucleoside HCV RNA polymerase inhibitor, following multiple, ascending, oral doses in patients with HCV genotype 1 infection who have failed prior interferon therapy [abstract]. *Annu Meet Am Assoc Study Liver Dis* 2007; 58: LB9.
59. Pharmasset, Inc. Pharmasset announces R7128 achieves 85% rapid virologic response in a 4-week combination study for the treatment of chronic hepatitis C (press release, January 7, 2008). Available at: http://www.pharmasset.com. Accessed July 2008.
60. Villano SA, Raible D, Harper D, et al. Antiviral activity of the non-nucleoside polymerase inhibitor, HCV-796, in combination with pegylated interferon Alfa-2b in treatment-naïve patients with chronic HCV [abstract]. *J Hepatol* 2007; 46(suppl 1): S24.
61. Rossignol JF, Elfert A, El-Gohary Y, Keefe EB. Randomized controlled trial of nitazoxanide-peginterferon-ribavirin, nitazoxanide-peginterferon and peginterferon-ribavirin in the treatment of patients with chronic hepatitis C genotype 4. *J Hepatol* 2008; 48(suppl 2): S30.
62. Rossignol JF, Kabil SM, El-Gohary Y, et al. Randomized double blind placebo-controlled trial of nitazoxanide in the treatment of patients with chronic hepatitis C genotype 4. *J Hepatol* 2008; 48(suppl 2): S311–S312.
63. Rossignol JF, Elfert A, Keeffe EB. Evaluation of a 4 week lead-in phase with nitazoxanide prior to nitazoxanide + peginterferon in treating chronic hepatitis C. *J Hepatol* 2008; 48(suppl 2): S312.
64. Flisiak R, Feinman SV, Jablkowski M, et al. Efficacy and safety of increasing doses of the cyclophilin inhibitor Debio 025 in combination with pegylated interferon alpha-2a in treatment naïve chronic HCV patients. *J Hepatol* 2008; 48(suppl 2): S62.
65. McHutchison JG, Bacon BR, Gordon SC, et al. Phase 1B, randomized, double-blind, dose-escalation trial of CPG 10101 in patients with chronic hepatitis C virus. *Hepatology* 2007; 46: 1341–1349.
66. Pockros P, Schiff E, Shiffman M, et al. The first caspase inhibitor of apoptosis, IDN-6556, given orally lowers liver enzymes in HCV patients [abstract]. *Hepatology* 2003; 38(suppl 1): 132A.
67. Encke J, Findeklee J, Geib J, et al. Prophylactic and therapeutic vaccination with dendritic cells against hepatitis C virus infection. *Clin Exp Immunol* 2005; 142: 362–369.
68. McHutchison JG, Patel K, Pockros P, et al. A phase I trial of an antisense inhibitor of hepatitis C virus (ISIS 14803), administered to chronic hepatitis C patients. *J Hepatol* 2006; 44: 88–96.
69. Stevenson M. Therapeutic potential of RNA interference. *N Engl J Med* 2004; 351: 1772–1777.
70. Poordad F, Shiffman M, Sherman K, et al. A study of telaprevir (TVR) with peginterferon Alfa-2a (P) and ribavirin (R) in subjects with well-documented prior P/R null response, non-response or relapse: preliminary results [abstract]. *J Hepatol* 2008; 48(suppl 2): LB.



don't publish online in the dark...

let our open linking network
shed some light.

This journal is CrossRef-ed! As a scientist or scholar navigating digital content, you've come to expect dynamic linking from citation to original source. As an author, you expect ready dissemination of your work to others. CrossRef is a collaborative, cross-publisher reference linking network that allows users to click on a citation and be taken directly to the full-text location. An independent, non-profit organization that employs open standards, CrossRef was established in cooperation with leading research publishers and the International DOI Foundation. Since the service went live in 2000, CrossRef has linked millions of articles from thousands of journals. Find out more at **www.crossref.org**, and make sure all your electronically published work is CrossRef-enabled.

crossref

CROSSREF.ORG | THE CITATION LINKING BACKBONE

200 WHEELER ROAD, 6TH FLOOR, BURLINGTON, MA 01803 • 781-359-2442

Fresh Content
Sliced Any Way You Like It

Wiley InterScience® OnlineBooks™

Content
Thousands of Wiley book titles across science, technology, medicine, business, and finance

Access
Single chapter or full title
One-time fee, subscription, or Pay-Per-View options

Capabilities
Search at the chapter level
Full text presented in PDF format
Citation download feature
COUNTER-compliant usage statistics

Satisfy your library's evolving research needs and budget with a customized blend of new content and licensing options at **Wiley InterScience® OnlineBooks™**. Pick and choose from an expanding roster of thousands of Wiley titles available online via **Wiley InterScience** to create a dynamic electronic book collection. Select from our new flexible licensing models — one-time fee, subscription, or Pay-Per-View. Add books as you need them from our distinguished portfolio of classic and cutting-edge titles. Whichever combination you select, enjoy unlimited concurrent access to content across the spectrum of science, medicine, technology, business, and finance.

WILEY
InterScience®
DISCOVER SOMETHING GREAT

WILEY
1807-2007 KNOWLEDGE FOR GENERATIONS

www.interscience.wiley.com

MOUNT SINAI JOURNAL OF MEDICINE

VOL. 75, NO. 5
SEPTEMBER/OCTOBER 2008

A Journal of Translational and Personalized Medicine

Issue Theme

Hospitalists: New Team in Forefront of Hospital Quality and Education

RECEIVED
OCT 22 2008

Defining Mission of Hospitalist Groups

Evolution and Future of Hospital Medicine

Economic and Healthcare Forces of Hospital Medicine

Hospitalist Educators: Future of Internal Medicine Training?

Redesigning Residency in Internal Medicine

Hospitalized Patient and Obesity

Pre-operative Evaluation of Hospitalized Patients

Palliative Care: DNR Does Not Mean DNRx

Special Features

Ethics of Complicated Medical Discharges: From "I'm Not Staying!" to "I'm not Leaving!"
Evidence-based Medicine: Clinical Prediction Rules for Hospitalists
Translational Medicine: Forging a New Cultural Identity

WILEY-BLACKWELL

Discover
Something
Great

Distinctive Content

Content
Journals • Backfile Collections
Reference Works & Encyclopedias
OnlineBooks™
Current Protocols Lab Manuals
Evidence-Based Resources

Access
Single article/chapter or full title
One-time fee, subscription,
or Pay-Per-View options
Unlimited concurrent usage

Capabilities
E-mail Alerts • EarlyView®
Citation download & tracking
Remote access • Mobile Edition®

Promote the process of scientific discovery with a different breed of research system: **Wiley InterScience®**. Long distinguished by its scope and extensive coverage of emerging disciplines, Wiley's authoritative content spans the spectrum of scientific, medical, technical, and business specialties. The dynamic **Wiley InterScience** platform adds value to your research by providing simple desktop access to this preeminent digital library of journals, reference works, books, laboratory manuals, evidence-based databases, and more. Researchers reap the combined benefits of top-tier resources and up-to-the-minute functionality. Experience the difference **Wiley InterScience** can make at your institution today.

WILEY
InterScience®
DISCOVER SOMETHING GREAT

WILEY
1807-2007 KNOWLEDGE FOR GENERATIONS

www.interscience.wiley.com

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

Editor-in-Chief

Penny A. Asbell

Deputy Editors

Dennis S. Charney
Maria New
C. Warren Olanow

Associate Editors

Stuart Aaronson
David H. Adams
Erwin Bottinger
Eric Hollander
Ravi Iyengar
Mary Klotman
Paul Klotman
Ihor R. Lemischka
Fred Lublin
Michael L. Marin
Thomas G. McGinn
Ira Nash
Rosamond Rhodes
Hugh Sampson
Mary Sano
Anthony Squire
Fred Suchy
Savio Woo

Editorial Board

William J. Bremner	University of Washington
Thomas M. Coffman	Duke University
Dwight Evans	University of Pennsylvania
Gabriel Haddad	University of California, San Diego
Heidi E. Hamm	Vanderbilt University
Matti Häyry	University of Manchester
Eve J. Higginbotham	Morehouse School of Medicine
Katherine A. High	University of Pennsylvania
Tony Hope	University of Oxford
John H. Krystal	Yale University
Raju Kucherlapati	Harvard Medical School
Robert J. Lefkowitz	Duke University
Robert Lisak	Wayne State University
George Lister	University of Texas
Philip A. Marsden	University of Toronto
Charles Nemeroff	Emory University
Scott Rauch	Harvard Medical School
Bernard Roizman	University of Chicago
Alan F. Schatzberg	Stanford University
Inder Verma	The Salk Institute
Jeffrey A. Whitsett	University of Cincinnati
Allan Wolkoff	Albert Einstein College of Medicine
Noam Zohar	Bar-Ilan University

Guest Editor

Andrew Dunn, MD, FACP
Associate Professor of Medicine
Director, Hospitalist Service
Mount Sinai Medical Center

Managing Editor

Stephen Cohen
stephen.cohen@mssm.edu
212-241-6304

Publisher

John Wiley & Sons
Anna Wesley, Associate Editor
awesley@wiley.com

Hospitalists: New Team in Forefront of Hospital Quality and Education

Themes for future issues of *MSJM*, each shown with Guest Editor(s)

- Advanced lung disease (Maria Padilla)
- Traumatic brain injury (Andrew Jagoda)
- Zoonotic infections (Stephanie Factor and Mary Klotman)
- Medical education (Lisa Coplit)
- Drug discovery in academic setting (Ravi Iyengar)
- Personalized medicine (Erwin Bottinger)

© 2008 **Mount Sinai School of Medicine.** All rights reserved. No part of this publication may be reproduced in any form or by any means, except as permitted under section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the publisher, or authorization through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, Phone: (978) 750-8400; Fax: (978) 750-4770. Such requests and other permission inquiries should be addressed to the Permissions Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; Phone: (201) 748-6011; Fax: (201) 748-6008; E-mail: permreq@wiley.com.

Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine (Print ISSN: 0027-2507; Online ISSN: 1931-7581 at Wiley InterScience, www.interscience.wiley.com) is published twelve times per year for the Mount Sinai School of Medicine by Wiley Subscription Services, a Wiley Company, 111 River Street, Hoboken, NJ 07030.

Send subscription inquiries to: Subscription Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-6645.

Offprint sales and inquiries should be addressed to the Reprint Billing Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-8776. All other inquiries should be addressed to the Customer Service Department at (201) 748-6645.

Individual subscription price: $95.00 in North America, $120.00 outside North America. Payment must be made in US dollars drawn on a US bank. Claims for undelivered copies will be accepted only after the following issue has been delivered. For more subscription prices and information, please visit www.mountsinaijournal.org.

Change of address: Please forward six weeks prior to move: present mailing label with change of address. Please enclose a copy of the mailing label or cite your subscriber reference number in order to expedite handling. Missing copies will be supplied when losses have been sustained in transit and where reserve stock permits. Send claims in care of: *Mount Sinai Journal of Medicine*, c/o John Wiley & Sons, Inc., Attn: Journals Admin Dept. UK, 111 River Street, Hoboken, NJ 07030.

Cancellations: Subscription cancellations will not be accepted after the first issue has been mailed. Periodicals postage paid at Hoboken, NJ and at additional mailing offices.

Postmaster: Send address changes to *Mount Sinai Journal of Medicine*, Subscription Distribution, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Printed in the United States of America.

Advertising inquiries should be addressed to: Advertising Sales Manager, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Phone: (201) 748-6921; Fax: (201) 748-6207. Advertising in this journal is published with a caveat emptor understanding. The Publisher, Editors, and members of the Editorial Board do not imply warranty of products, nor quality, validity, or approval of education material offered by such advertising.

Indexed by: Biological Abstracts® (Thomson ISI), Chemical Abstracts Service/SciFinder (ACS), Current Contents®/Clinical Medicine (Thomson ISI), EMBASE/Excerpta Medica (Elsevier), Index Medicus/MEDLINE/PubMed (NLM), Science Citation Index Expanded™ (Thomson ISI), SCOPUS (Elsevier), Web of Science® (Thomson ISI).

This journal is printed on acid-free paper.

Disclaimer: All articles published, including but not limited to original research, reviews, clinical notes, clinical pearls, prismatic cases, clinician's corner, editorials, reports, and letters, represent the opinions and views of the authors and do not reflect any official policy or medical opinion of the Mount Sinai School of Medicine, the institutions with which the authors are affiliated, or the Publisher unless this is clearly specified. Articles published herein should not be relied upon as recommending or promoting a specific method, diagnosis, or treatment by physicians for any particular patient or disorder. Although the Editors and Publisher believe that drug selections and dosages and the specifications and usage of equipment and devices as set forth herein are in accord with current recommendations and practice at the time of publication, they accept no legal responsibility for any errors or omissions, and make no warranty, express or implied, with respect to material contained herein.

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

CONTENTS
VOL 75, NO 5

Theme Introduction

415 Hospital Medicine in High-Def: Importance of Defining Local Mission of Hospitalist Groups
Andrew Dunn

Theme Articles

418 Evolution and Future of Hospital Medicine
Kevin J. O'Leary and Mark V. Williams

424 Economic and Healthcare Forces of Hospitalist Movement
Tara F. Bishop and Navneet Kathuria

430 Hospitalist Educators: Future of Internal Medicine Training?
Jill Goldenberg, Jeffrey J. Glasheen

436 Achieving Promise of Hospital Medicine Through Internal Medicine Residency Redesign
Jeffrey J. Glasheen, Jill Goldenberg, and John R. Nelson

442 Pre-operative Evaluation of Hospitalized Patients
Stephen K Sigworth

449 Prevention of Venous Thromobembolism in Hospitalized Medical Patients
Joshua D. Lenchus and Amir K. Jaffer

454 Obesity and Hospitalized Patient
Brian Markoff and Alison Amsterdam

460 DNR Does Not Mean DNRx: How Palliative Care Can Help Facilitate Communication about Goals of Care in Advanced Illness
Cardinale B. Smith and Lynn Bunch

Feature Departments

466 Ethics (Editor: Rosamond Rhodes)
From "I'm Not Staying!" to "I'm Not Leaving!"–Ethics, Communication, and Empathy in Complicated Medical Discharges
David Alfandre

472 Evidence-based Medicine (Editor: Thomas McGinn)
Clinical Prediction Rules for the Hospitalist
Ramiro Jervis and Thomas McGinn

Special Feature

478 Translational Medicine: Forging a New Cultural Identity
Barry Coller

THEME INTRODUCTION

Hospital Medicine in High-Def: Importance of Defining Mission of Hospitalist Groups

Andrew Dunn, MD, FACP

Division of General Internal Medicine, Mount Sinai School of Medicine, New York, NY

Leaders in the specialty of hospital medicine have worked to define the key roles that hospitalists can play in improving the care of hospitalized patients.[1,2] This issue of the *Mount Sinai Journal of Medicine* is dedicated to the variety of services that have come to be provided by hospitalists, including clinical care of medical patients, medical consultation and comanagement, perioperative assessment and management, education, expertise in navigating and improving hospital systems, and contributions to the bottom line (the business of medicine).[3–6] However, having each individual hospitalist group define its core mission within its hospital is equally as essential as national leaders defining roles for hospitalists across the country. Defining these core goals will be beneficial for all parties: the individual hospitalist group members, the hospitalist service, and the hospital.

The expression "if you have seen 1 hospital medicine program, you have seen 1 hospital medicine program" concisely portrays the heterogeneity of hospitalist services across the country. Hospitalist programs may be part of community hospitals, university hospitals, or teaching hospital affiliates; may be a major source of referrals for private physicians or care only for "service" populations; may have a major role in education and/or hospital quality initiatives or solely provide clinical care; and may or may not include prominent roles in medical consultation, preoperative clinics, comanagement, geriatrics, palliative care, and other clinical services. Given the vast array of possible expectations, the potential for confusion and uncertainty will be high unless each program defines its mission within its institution. Defining program goals can help:

- Promote a sense of purpose by team members in understanding new initiatives in the context of striving to meet and achieve goals.
- Identify characteristics sought in future group members during the interview process.
- Determine time, staff, and other resources needed to fulfill the core roles.

SENSE OF TEAM PURPOSE

We are currently living in an era of intense expansion of hospitalist roles and services. The rapid growth often results in hospitalist groups being asked to take on sudden or unexpected roles, such as expansion of nonteaching services and comanagement of surgical services. Aggressive expansion, particularly into new product lines, is often a source of anxiety for group members. Members may feel a sense of uncertainty about the duties that they will be asked to perform or the caseload that they will be expected to carry. By conceptualizing growth as aligning with the group's mission, team members can see expansion as a recognition of their expertise and accomplishments, rather than a threat to their current status. Hospitalist growth dictated by hospital needs that does not align with the mission of the service also needs to be addressed. One strategy for avoiding these situations is to have frequent discussions with hospital and

Address Correspondence to:

Andrew Dunn, MD, FACP
Division of General Internal Medicine
Mount Sinai School of Medicine
New York, NY
Email: andrew.dunn@mountsinai.org

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20068

© 2008 Mount Sinai School of Medicine

departmental leaders concerning the focus of the service and the areas in which the group will be of best use to the hospital. If a responsibility that does not align perfectly with the group's goals is unavoidable, reconceptualizing and/or reshaping the role to fit the framework that the group has developed may make the role more acceptable. For example, in-house night coverage can be considered an important initiative in enhancing quality and patient safety or might have a teaching component if oversight of house staff is part of the new role. Lastly, when the group takes on a responsibility that clearly does not fit the mission of the service, members will need to understand that these duties are essential for the hospital, although apart from the core services provided by the group. If it is necessary, these tasks can be compartmentalized, so that group members who are not interested in a career path in hospital medicine (eg, those transitioning to a fellowship) are preferentially assigned the noncore roles and the career-oriented hospitalist can focus on the mission of the service.

INTERVIEWING AND HIRING

Hospitalist applicants are as heterogeneous as hospitalist programs. Different candidates have different training, backgrounds, personal characteristics, and, most importantly, professional preferences. An explicit and overarching mission statement will help clarify the desired attributes of potential team members. For example,

- A service developing its niche as the main educator for an internal medicine residency program may target applicants who have a desire to teach and develop curricula.
- A program whose goal is to be involved in hospital systems may want applicants with backgrounds and interests in improving hospital processes.
- A program seeking to expand clinical care throughout the hospital may want to recruit physicians with strong clinical skills who are interested in medical consultation or comanagement.

Recruiting and hiring candidates whose personal goals align with the goals of the service will maximize the career satisfaction of team members. This will help foster a healthy working environment for the service as well as productivity and retention.

TIME AND RESOURCES

Clearly defining the group's roles is essential to ensure that the appropriate resources are provided. Most importantly, adequate time needs to be allotted for the various responsibilities to be performed. For example, if the hospitalists are expected to play important roles on hospital committees, such as those involving patient safety initiatives (eg, reducing the incidence of *Clostridium difficile* colitis or meeting goals for core measures), or if there is an expectation that the hospitalists will be the primary teachers for aspects of the residency program, the hospital will need to recognize these crucial roles and ensure that an adequate amount of nonclinical time is allotted. Expecting hospitalists to provide uninterrupted clinical care while contributing to these other essential hospital functions will lead to dissatisfaction, burnout, turnover, and poor recruitment.

A challenge for each hospitalist program is to create a hospital metric or approach that accounts for the contribution to quality initiatives, the educational mission, or other roles. For certain activities, the expected hours of work can be strictly calculated; major teaching roles at university hospitals or administrative positions requiring substantial amounts of time are often handled in this manner. For many hospitalist tasks, however, recognition of the essential nature of the role and the fact that there is a substantial time commitment required to fulfill the role may suffice. For example, the hospital may want a hospitalist to lead an initiative to decrease the risk of central venous line infections or meet the requirements for the Joint Commission's National Patient Safety Goals. The precise amount of time allotted away from direct patient care will depend on multiple factors, including the scope of the responsibilities and local factors, such as financial considerations and culture. A priority for the Society of Hospital Medicine and other leaders in hospital medicine should be to promote a standard approach to this complex issue in order to minimize local factors, promote greater uniformity across hospitals, and ensure that the necessary time for these tasks is provided.

CONCLUSION

The trend for hospitals to rely on hospitalists to address many of their essential needs aligns well with the interests of hospitalists, many of whom have a passion for improving hospital systems and patient outcomes. However, it will be important for each local hospitalist group to consider its strengths, goals, and mission, instead of accumulating an increasing number of diverse responsibilities without reflection of how new roles fit into the group's current responsibilities. An expansion of those roles that align with the group's strengths can be sought, and responsibilities that do not conform to the mission can be avoided or adopted in an explicit manner that minimizes disruption to the group. Defining the goals will allow the recruitment of hospitalists who can best contribute to the team's success. Clearly defining the key roles played by the hospitalist service will also allow for explicit discussion of the need for appropriate time and resources to be allotted to the group to meet expectations. Providing adequate support to hospitalists to fulfill these responsibilities will improve recruitment and retention, allow the service to develop stability and continuity, and provide the necessary environment for the service to meet hospital goals and deliver outstanding clinical outcomes.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Wachter RM. The state of hospital medicine in 2008. *Med Clin North Am* 2008; 92: 265–273.
2. Amin A, Ansari A, Blatt MB, et al. The hospitalist model of care: a positive influence on efficiency, quality of care, and outcomes. *Crit Path Cardiol* 2004; (suppl): S5–S7.
3. Saint S, Flanders SA. Hospitalists in teaching hospitals: opportunities but not without danger. *J Gen Intern Med* 2004; 19: 392–393.
4. Auerbach AD, Wachter RM, Katz P, et al. Implementation of a voluntary hospitalist service at a community teaching hospital: improved clinical efficiency and patient outcomes. *Ann Intern Med* 2002; 137: 859–865.
5. Huddleston JM, Long KH, Naessens JM, et al. Medical and surgical comanagement after elective hip and knee arthroplasty: a randomized controlled trial. *Ann Intern Med* 2004; 141: 28–38.
6. Meltzer D, Manning WG, Morrison J, et al. Effects of physician experience on costs and outcomes on an academic general medicine service: results of a trial of hospitalists. *Ann Intern Med* 2002; 137: 866–874.

The Evolution and Future of Hospital Medicine

Kevin J. O'Leary, MD, and Mark V. Williams, MD

[1]Division of Hospital Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL

ABSTRACT

Hospitalists practice hospital medicine, which represents the fastest growing medical speciality in the history of American medicine. Initially fueled by research demonstrating reductions in lengths of stay and hospital costs, the hospitalist movement has continued to gather momentum as a response to residency work hour revisions, patient flow challenges in hospitals, and a growing national focus on patient safety. Hospitalists have expanded the clinical services they provide to include coverage of the critically ill, palliative care, and surgical co-management. Hospitalists increasingly participate in nonclinical activities as well, including teaching, research, quality improvement, and clinical informatics.

Challenges remain for this new field of medicine and present opportunities for research and innovation. Work life satisfaction may be enhanced by efforts to optimize workload and redesign care delivery models. Hospitalists may continue to demonstrate value to stakeholders by leading efforts to improve quality of care, especially in areas such as interdisciplinary communication and smoothing patients' transitions between care settings. Hospitalists should continue to take an active role in improving clinical efficiency and work with hospitals to fully leverage the capabilities of information and telecommunication systems. Looking to the future, hospitalists have exciting potential for a positive impact in a broad range of areas and lasting effect on healthcare delivery. ***Mt Sinai J Med 75:418–423, 2008.*** *© 2008 Mount Sinai School of Medicine*

Address Correspondence to:

Kevin J. O'Leary
Division of Hospital Medicine
Northwestern University
Feinberg School of Medicine
Chicago, IL
Email: koleary@nmff.org

Key Words: hospitalists, hospital medicine, healthcare systems, medical history.

Hospitalists are physicians whose primary professional focus is the general medical care of hospitalized patients.[1] Hospitalists practice hospital medicine, which represents the fastest growing medical specialty in the history of American medicine. There were approximately 1000 internal medicine physicians focusing their practice predominantly in hospitals in the mid 1990s when Wachter and Goldman first coined the term *hospitalist* in 1996 to describe this nascent model of care.[2,3] Since then, the movement has experienced intense growth, with about 23,000 hospitalists in 2008 and with expectations of more than 30,000 in the next few years. Over one-third of US hospitals had practicing hospitalists in 2004,[4] and we suspect that half of hospitals now have them. Figure 1 shows the number of hospitalists in the United States with respect to other medical specialties, demonstrating that there will soon be more hospitalists than cardiologists.[3,5]

The initial forces driving the spread of hospitalists across the United States included cost pressures on hospitals, physician groups, and managed care organizations; the increased acuity of hospitalized patients; and the time pressures on primary care physicians in the office.[6] With the prospect of shorter lengths of stay and lower costs, hospitals implemented hospital medicine programs in an effort to improve their financial performance. A number of studies have now confirmed that hospitalists do indeed reduce lengths of stay and costs.[7–12] Evidence-based reviews have found average reductions of 12% to 17% in lengths of stay and 13% to 14% in costs.[13,14] Overall, patients appear satisfied with hospitalist care.[13] One study found that, although most patients agreed that care from a physician with whom they have had a long-term relationship is beneficial, they also agreed with the importance of being cared for by a physician skilled in treating hospitalized patients and available throughout the day.[15]

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20078

© 2008 Mount Sinai School of Medicine

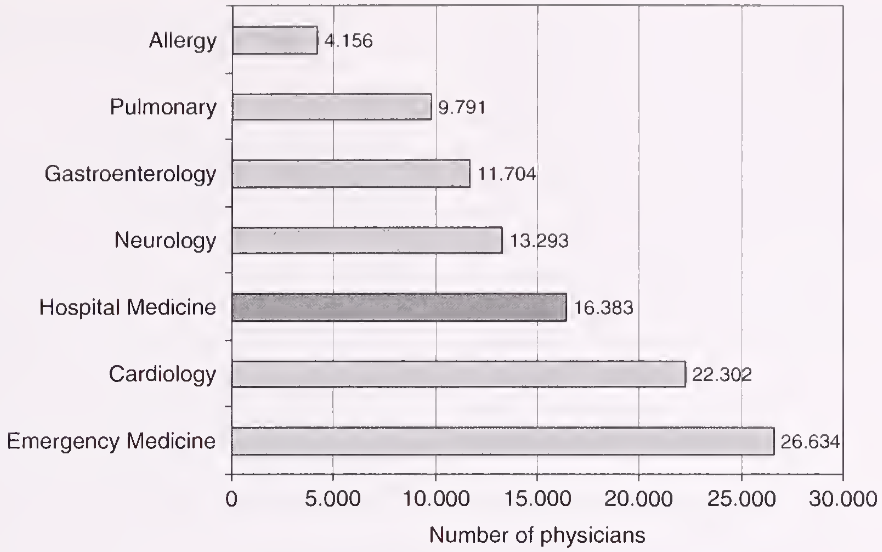


Fig 1. The number of hospitalists in the United States with respect to other medical specialties. Adapted from *Physician Characteristics and Distribution in the US, 2005 Edition.*[5]

The question of whether hospitalists improve patient outcomes remains unsettled. Although 2 studies have demonstrated a reduction in mortality with the use of hospitalists,[7,12] other studies have not.[16,17] Importantly, no studies have revealed worsened outcomes from hospitalist care as many feared would be a consequence of increased efficiency and the resultant reduced length of hospital stay. Although hospitalists were initially embraced by community hospitals, revisions to resident work-hour standards[18] rapidly drove the development and expansion of clinical programs at academic medical centers.[19] Additional drivers of hospitalist program growth have included patient-flow challenges in hospitals, rising malpractice costs, and the growing national focus on patient safety.[20]

VALUE ADDED BY HOSPITALISTS

The traditional business case for the use of hospitalists posits that reductions in lengths of stay and costs per case result in sufficient savings for the hospital to fund program costs not offset by collections from billings. Yet hospital medicine leaders have wondered, as reduced lengths of stay and lower costs plateau, how they will continue to demonstrate their value to the hospitals that help support them. Hospitalists and the hospitals with which they work have identified opportunities that arise from having a group of physicians housed in the hospital and focused on clinical care delivery there (Table 1).

The expansion of existing clinical services and the addition of new ones are the most straightforward ways for hospital medicine leaders to collaborate with their hospitals if a sufficient number of hospitalists are available. The majority of hospital medicine groups now comanage surgical patients and provide coverage for critical care patients.[21] Most hospital medicine groups also cover unassigned patient admissions from the Emergency Department, and many cover cardiac arrests. Additionally, some hospitalists have expanded into setting up palliative care

Table 1. *Value Added by Hospitalists.*

Clinical activities
General medical care of hospitalized patients Medical consultations Comanagement of surgical patients Critical care coverage Care for unassigned patients from the Emergency Department Cardiac arrest and rapid response team coverage Procedures (eg, thoracentesis and paracentesis) Palliative care
Nonclinical activities
Committee service Quality improvement and patient safety Pharmacy and therapeutics Utilization review Informatics/information systems Teaching Research

consultation services.[22] Nonclinical activities by hospitalists offer potentially the greatest value to hospital administrators through the provision of committed physicians seeking to improve the overall system of care delivery. With declining interest and availability among outpatient-based physicians, participation and leadership by hospitalists in committees (eg, utilization review, pharmacy and therapeutics, and quality council) provide clear opportunities for substantive contributions. Increasingly, hospitals also rely on hospitalists to help implement electronic health records and computerized provider order entry. Approximately half of all hospital medicine programs report collaborating with their hospital information systems departments.[21]

As already noted, many academic medical centers started hospital medicine programs in an effort to help residency programs comply with revised Accreditation Council for Graduate Medical Education work-hour guidelines. Given the increased emphasis on ambulatory care training in internal medicine[23] and the possibility that residency work hours may some day be further reduced, the demand for hospitalists in academic medical centers will likely increase. Beyond the benefit in work-hour compliance, hospitalists have shown their value to academic medical centers as outstanding teachers. In 1 study, medical students and residents gave higher ratings to hospitalists and general internists than their specialist counterparts.[24] Another study found that hospitalists surpassed all nonhospitalist attendings in their teaching evaluations.[25] Initial concerns that hospitalist clinician educators would adversely affect resident autonomy and decision-making abilities[26] have been largely disproved.[27,28]

Hospitalists have also sought to add value to their hospitals by collaborating on (or leading) quality improvement and patient safety efforts. Hospitalists are well suited for this type of work because they have the necessary clinical expertise and a vested interest in improving the systems in which they work. In fact, the majority of hospitalists participate in quality improvement efforts at their hospitals.[21] Many have also been instrumental in implementing rapid response teams and systems for medication reconciliation.[21] For hospitalists to evolve from participating in quality improvement initiatives to leading them, they will need to obtain formal training such as the Advanced Training Program in Health Care Delivery Improvement (http://intermountainhealthcare.org/xp/public/institute/courses/atp) offered by Intermountain Healthcare in Utah or Six Sigma training available throughout the industry.

EMERGENCE OF NEW FIELD: HOSPITAL MEDICINE

Hospital medicine has emerged as a new field in medicine, following a course similar to that of other site-based specialties such as emergency medicine and critical care medicine. Several textbooks on hospital medicine exist. The *Journal of Hospital Medicine*, first published in January 2006, receives over 300 submissions per year and is already included in the National Library of Medicine's Medical Literature Analysis and Retrieval System Online.[29] A number of residency training tracks and numerous fellowships in hospital medicine are available.[30] The Society of Hospital Medicine, highly regarded among medical professional societies, is now over 6000 members strong. In 2006, the Society of Hospital Medicine published core competencies in hospital medicine to standardize the expectations of practicing hospitalists and serve as a foundation for curricula and professional development for hospitalists.[31]

The American Board of Internal Medicine has worked closely with the Society of Hospital Medicine to determine whether separate certification might be the logical next step for the field of hospital medicine. Goals of certification include demonstration of ongoing competence and promotion of lifelong learning. Most hospitalists are trained as generalists, entering the field after completion of an internal medicine, pediatric, or, less often, family practice residency. Separate certification would necessarily lead to a reassessment of these residency training programs, prompting revisions and possibly prolonging training requirements. In the absence of evidence suggesting a need for revisions to residency training programs, the American Board of Internal Medicine endorsed the concept of recognition of focused practice in hospital medicine.[32] Currently, the American Board of Medical Specialties is evaluating the recommendation of the American Board of Internal Medicine and should provide a final decision by the end of 2008. If it is approved, board-certified graduates of internal medicine residency will be able to obtain focus recognition when they recertify in internal medicine.

REMAINING CHALLENGES

The continued success of the hospital medicine movement will require careful attention to remaining challenges (Table 2). Although 1 study found that 90% of hospitalists were satisfied or very satisfied with their careers, the same study reported that 38% were either burnt out or at risk for burnout.[33] In a

Table 2. *Hospitalist Challenges and Corresponding Opportunities.*

Challenges	Opportunities
Work-life satisfaction/fulfillment	Research in staffing/coverage models and workload
Communication between hospitalists and outpatient physicians	Leverage of information and telecommunication systems
Interdisciplinary communication and teamwork	Research in team structure and communication tools
Adverse events that occur at hospital discharge	Multifaceted evaluations and redesign of discharge process
Requested expansion of clinical services	Improved clinical efficiency for additional clinical services

national survey, hospitalist leaders listed workload and work-life balance among their top concerns.[21] Although research provides some clarity in this area,[34,35] the optimal workload and care delivery model for hospitalists has yet to be determined.

The rising demand for hospitalists is currently outpacing the supply. The increased competition for hospitalists means that many hospital medicine programs are continuously recruiting, and hospitalist salaries have risen 13% in the past 2 years.[36] The rapid expansion of hospital medicine also translates into a relatively young work force, as open positions are often filled by recent graduates from residency. The mean age is 37 years, and hospitalists have been practicing for a mean of 3.7 years.[36]

Communication is critical in hospital medicine. Although studies show that the majority of primary care physicians are satisfied with the overall care delivered by hospitalists,[37,38] they are less well satisfied with hospitalist communication.[39] Because expectations and preferences vary at the hospital level, hospitalist leaders should regularly meet with outpatient physicians to clarify optimal local communication practices.

Communication with other members of the care delivery team within the hospital is also critical. Although teamwork and collaboration have been well studied in operating rooms and intensive care units,[40–42] relatively little research exists for the general medical inpatient setting. One study found that nurses and physicians discussed patients' plans of care with one another only 50% of the time and agreed on the priorities of care in a mere 17% of cases.[43] Future research should better characterize communication patterns and define barriers to communication between hospitalists and other inpatient healthcare team members. With this knowledge, methods to improve interdisciplinary communication and teamwork can be designed.

Before the very first hospitalists, patients and their caregivers perceived major problems related to the hospital discharge process.[44] Studies have revealed that preventable adverse events occur in approximately 12% of patients in the period immediately following hospital discharge,[45,46] Patients are often released from the hospital with test results still pending[47] and clinical issues that need further evaluation in the outpatient setting. Hospitalists recognize the need to prioritize the discharge process as an area in dire need of improvement.[48] Existing tools to improve safety at discharge have been disseminated,[49] and a call has been made for the discovery of additional methods to improve safety.[48] Of note, the John A. Hartford Foundation provided a $1.4 million grant to the Society of Hospital Medicine to develop a toolkit to improve the discharge process and disseminate this to hospitals nationwide.

OPPORTUNITIES FOR THE FUTURE

Certainly, the aforementioned challenges present great opportunities for research and innovation. Two other areas, already mentioned as value-added services provided by hospitalists, will continue to grow as opportunities in the years to come. The first is information systems. In light of the fact that only a small fraction of hospitals have fully implemented computerized provider order entry,[50] hospitalists can anticipate an opportunity to provide value in informatics for the foreseeable future. Hospitalists should seize the opportunity to leverage information systems to improve communication with outpatient physicians. Similarly, communication between inpatient health care team members might be enhanced through innovations in information and/or telecommunication systems.

The second area in which hospitalists have an opportunity to add value is quality improvement and patient safety. As the Center for Medicare and Medicaid Services and other payers begin to hold hospitals accountable for their performance in quality, the role of the hospitalist in quality improvement will expand. Beginning in October 2008, the Center for Medicare and Medicaid Services will no longer pay hospitals for certain conditions that were not present on admission. In addition, the Center for Medicare and Medicaid Services will reimburse hospitals differentially on the basis of their performance in evidence-based processes of care. In the era of pay for performance, hospitals will instinctively look to hospitalists as their partners in improving quality and safety performance.

One note of caution is warranted in an examination of the future of hospital medicine. In the coming years, hospitalist leaders will face an increasing need to demonstrate value to their stakeholders in new ways. Undoubtedly, these leaders will be asked to provide care for new clinical services, some of which may not be considered to be within the traditional scope of practice for generalists. Hospitalist leaders must be careful that the expansion of hospitalists into new service lines makes sense for the patients and occurs at a pace reasonable for their colleagues and staff.

CONCLUSION

The ascension of the hospitalist movement culminated in the development of a new field of medicine: hospital medicine. A product of a convergence of healthcare forces, this new field will continue to grow in light of ongoing trends in healthcare. For the field to achieve its full potential, certain challenges will need to be addressed as opportunities for research and innovation. Looking to the future, hospitalists have exciting potential for a positive impact in a broad range of areas and a lasting effect on healthcare delivery in hospitals and beyond.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Society of Hospital Medicine. Frequently asked questions: What is a hospitalist? Available at: http://www.hospitalmedicine.org/content/navigation-menu/aboutshm/generalinformation/general_information.htm. Accessed February 2008.
2. Wachter RM, Goldman L. The emerging role of "hospitalists" in the American health care system. *N Engl J Med* 1996; 335: 514–517.
3. Lurie JD, Miller DP, Lindenauer PK, et al. The potential size of the hospitalist workforce in the United States. *Am J Med* 1999; 106: 441–445.
4. Kralovec PD, Miller JA, Wellikson L, Huddleston JM. The status of hospital medicine groups in the United States. *J Hosp Med* 2006; 1: 75–80.
5. American Medical Association. *Physician Characteristics and Distribution in the US*, 2005 Edition. Chicago, IL.
6. Williams MV. The future of hospital medicine: evolution or revolution? *Am J Med* 2004; 117: 446–450.
7. Auerbach AD, Wachter RM, Katz P, et al. Implementation of a voluntary hospitalist service at a community teaching hospital: improved clinical efficiency and patient outcomes. *Ann Intern Med* 2002; 137: 859–865.
8. Halpert AP, Pearson SD, LeWine HE, McKean SC. The impact of an inpatient physician program on quality, utilization, and satisfaction. *Am J Manag Care* 2000; 6: 549–555.
9. Kaboli PJ, Barnett MJ, Rosenthal GE. Associations with reduced length of stay and costs on an academic hospitalist service. *Am J Manag Care* 2004; 10: 561–568.
10. Palmer HC, Armistead NS, Elnicki DM, et al. The effect of a hospitalist service with nurse discharge planner on patient care in an academic teaching hospital. *Am J Med* 2001; 111: 627–632.
11. Wachter RM, Katz P, Showstack J, et al. Reorganizing an academic medical service: impact on cost, quality, patient satisfaction, and education. *JAMA* 1998; 279: 1560–1565.
12. Meltzer D, Manning WG, Morrison J, et al. Effects of physician experience on costs and outcomes on an academic general medicine service: results of a trial of hospitalists. *Ann Intern Med* 2002; 137: 866–874.
13. Coffman J, Rundall TG. The impact of hospitalists on the cost and quality of inpatient care in the United States: a research synthesis. *Med Care Res Rev* 2005; 62: 379–406.
14. Wachter RM, Goldman L. The hospitalist movement 5 years later. *JAMA* 2002; 287: 487–494.
15. Hruby M, Pantilat SZ, Lo B. How do patients view the role of the primary care physician in inpatient care? *Am J Med* 2001; 111: 21S–25S.
16. Lindenauer PK, Rothberg MB, Pekow PS, et al. Outcomes of care by hospitalists, general internists, and family physicians. *N Engl J Med* 2007; 357: 2589–2600.
17. Southern WN, Berger MA, Bellin EY, et al. Hospitalist care and length of stay in patients requiring complex discharge planning and close clinical monitoring. *Arch Intern Med* 2007; 167: 1869–1874.
18. Philibert I, Friedmann P, Williams WT. New requirements for resident duty hours. *JAMA* 2002; 288: 1112–1114.
19. Saint S, Flanders SA. Hospitalists in teaching hospitals: opportunities but not without danger. *J Gen Intern Med* 2004; 19: 392–393.
20. Pham HH, Devers KJ, Kuo S, Berenson R. Health care market trends and the evolution of hospitalist use and roles. *J Gen Intern Med* 2005; 20: 101–107.
21. Society of Hospital Medicine. 2005–2006 survey. Available at: http://www.hospitalmedicine.org//am/template.cfm?section=home. Accessed February 2008.
22. Pantilat SZ. Palliative care and hospitalists: a partnership for hope. *J Hosp Med* 2006; 1: 5–6.
23. Meyers FJ, Weinberger SE, Fitzgibbons JP, et al. Redesigning residency training in internal medicine: the consensus report of the Alliance for Academic Internal Medicine Education Redesign Task Force. *Acad Med* 2007; 82: 1211–1219.
24. Kripalani S, Pope AC, Rask K, et al. Hospitalists as teachers. *J Gen Intern Med* 2004; 19: 8–15.
25. Hauer KE, Wachter RM, McCulloch CE, et al. Effects of hospitalist attending physicians on trainee satisfaction with teaching and with internal medicine rotations. *Arch Intern Med* 2004; 164: 1866–1871.

26. Kemper AR, Freed GL. Hospitalists and residency medical education: measured improvement. *Arch Pediatr Adolesc Med* 2002; 156: 858–859.
27. Kulaga ME, Charney P, O'Mahony SP, et al. The positive impact of initiation of hospitalist clinician educators. *J Gen Intern Med* 2004; 19: 293–301.
28. Landrigan CP, Muret-Wagstaff S, Chiang VW, et al. Senior resident autonomy in a pediatric hospitalist system. *Arch Pediatr Adolesc Med* 2003; 157: 206–207.
29. Williams MV. Two years done and more evolution. *J Hosp Med* 2008; 3: 4–5.
30. Society of Hospital Medicine. Hospitalist fellowships. Available at: http://www.hospitalmedicine.org/content/navigationmenu/education/hospitalmedicineprograms/hospital_medicine_pr.htm. Accessed February 2008.
31. Dressler DD, Pistoria MJ, Budnitz TL, et al. Core competencies in hospital medicine: development and methodology. *J Hosp Med* 2006; 1(suppl 1): 48–56.
32. Wachter RM. What will board certification be-and mean-for hospitalists? *J Hosp Med* 2007; 2: 102–104.
33. Hoff TH, Whitcomb WF, Williams K, et al. Characteristics and work experiences of hospitalists in the United States. *Arch Intern Med* 2001; 161: 851–858.
34. O'Leary KJ, Lindquist L, Colone MA, et al. Effect of a hospitalist-care coordinator team on a nonteaching hospitalist service. *J Hosp Med* 2008; 3: 103–109.
35. O'Leary KJ, Liebovitz DM, Baker DW. How hospitalists spend their time: insights on efficiency and safety. *J Hosp Med* 2006; 1: 88–93.
36. Society of Hospital Medicine. 2007–2008 survey. Available at: http://www.hospitalmedicine.org/am/template.cfm?section=surveys2&template=/cm/htmldisplay.cfm&contentid=18412. Accessed July 2008.
37. Auerbach AD, Nelson EA, Lindenauer PK, et al. Physician attitudes toward and prevalence of the hospitalist model of care: results of a national survey. *Am J Med* 2000; 109: 648–653.
38. Fernandez A, Grumbach K, Goitein L, et al. Friend or foe? How primary care physicians perceive hospitalists. *Arch Intern Med* 2000; 160: 2902–2908.
39. Pantilat SZ, Lindenauer PK, Katz PP, Wachter RM. Primary care physician attitudes regarding communication with hospitalists. *Am J Med* 2001; 111: 15S–20S.
40. Lingard L, Espin S, Whyte S, et al. Communication failures in the operating room: an observational classification of recurrent types and effects. *Qual Saf Health Care* 2004; 13: 330–334.
41. Makary MA, Sexton JB, Freischlag JA, et al. Operating room teamwork among physicians and nurses: teamwork in the eye of the beholder. *J Am Coll Surg* 2006; 202: 746–752.
42. Thomas EJ, Sexton JB, Helmreich RL. Discrepant attitudes about teamwork among critical care nurses and physicians. *Crit Care Med* 2003; 31: 956–959.
43. Evanoff B, Potter P, Wolf L, et al. Can we talk? Priorities for patient care differed among health care providers. Advances in Patient Safety: From Research to Implementation. Volumes 1-4, AHRQ Publication Nos. 050021 (1–4). http://www.ahrg.gov/qual/advances/. Accessed March 3, 2008.
44. Cleary PD, Edgman-Levitan S, Roberts M, et al. Patients evaluate their hospital care: a national survey. *Health Aff (Millwood)* 1991; 10: 254–267.
45. Forster AJ, Clark HD, Menard A, et al. Adverse events among medical patients after discharge from hospital. *CMAJ* 2004; 170: 345–349.
46. Forster AJ, Murff HJ, Peterson JF, et al. The incidence and severity of adverse events affecting patients after discharge from the hospital. *Ann Intern Med* 2003; 138: 161–167.
47. Roy CL, Poon EG, Karson AS, et al. Patient safety concerns arising from test results that return after hospital discharge. *Ann Intern Med* 2005; 143: 121–128.
48. Coleman EA, Williams MV. Executing high-quality care transitions: a call to do it right. *J Hosp Med* 2007; 2: 287–290.
49. Halasyamani L, Kripalani S, Coleman E, et al. Transition of care for hospitalized elderly patients–development of a discharge checklist for hospitalists. *J Hosp Med* 2006; 1: 354–360.
50. Jha AK, Ferris TG, Donelan K, et al. How common are electronic health records in the United States? A summary of the evidence. *Health Aff (Millwood)* 2006; 25: 496–507.

Mount Sinai Journal of Medicine 75:424–429, 2008

Economic and Healthcare Forces of Hospitalist Movement

Tara F. Bishop, MD and Navneet Kathuria, MD, MPH

Mount Sinai School of Medicine, New York, NY

ABSTRACT

The field of hospital medicine has become a widely accepted model for inpatient care and has grown rapidly in the past ten years. The impetus for growth has largely been pressure to contain costs for inpatient care and improve efficiency in the hospital. Studies have shown that care by hospitalists is generally more cost-effective than care by faculty or private practice physicians without affecting quality. The field faces challenges in continuity of patient care and retention of physicians in the workforce. ***Mt Sinai J Med 75:424–429, 2008.*** © *2008 Mount Sinai School of Medicine*

Key Words: Hospital Costs, Hospitalist Economics, Hospitalist Workforce, Hospitalist Cost-Effectiveness, Hospitalist Quality.

In an article in the *New England Journal of Medicine* in 1996,[1] Wachter and Goldman described a new group of physicians whose practice was hospital-based. *Hospitalists*, as Wachter and Goldman called them, were internists whose majority of patient care was in the inpatient arena.

Since Wachter and Goldman's initial publication,[1] hospitalist medicine has been one of the fastest growing fields in medicine. By 1999, 28% of internists reported using hospitalists for some inpatient care.[2] Today, there are more than 12,000 hospitalists practicing in the United States,[3] and some have projected that the field will grow to 30,000 physicians, that is, as many as in the field of cardiology.[4]

The impetus for growth has been a combination of financial pressure on hospitals and health insurance plans and time pressure on outpatient-based primary care physicians (PCPs). Other forces that have contributed to the rise in hospitalists include efforts to improve patient safety and hospital flow.[5] The literature reflects these trends. Most studies have looked at the cost benefits of using hospitalists and the quality of patient care by hospitalists. The field has continued to grow because the literature has consistently shown that hospitalists lower hospital costs and shorten length of stay while keeping the quality of patient care high. Other issues have emerged such as transitions in care and the sustainability of the workforce. This article reviews the financial pressure that led to the growth in hospitalists, the cost effectiveness and quality data, and the challenges that the field faces in the future.

Address Correspondence to:

Navneet Kathuria
Mount Sinai School of Medicine
Email: navneet.kathuria@
exchange.mssm.edu

HEALTHCARE ENVIRONMENT

The major reason that hospitalist medicine emerged in the mid 1990s was cost pressure. With the growth in managed care, PCPs were required to be more efficient, and efficiency was defined by the number of patients seen. Commuting to the hospital from the office was a barrier to efficiency, particularly if the PCP was traveling long distances to see fewer inpatients.

Hospitals were also forced to reduce the cost of care and length of stay while improving the flow of patients and the quality of clinical care. By 1996, several hospitals in California were trying a model in which hospitalist-based physicians–hospitalists–were used to control hospital costs and relieve PCPs of the burden of inpatient care.[1]

More recent trends in health care have supported the growth of the hospitalist movement. Pham *et al.*[6] found that the growth in hospitalists continues to be fueled by financial pressure on hospitals and time pressure on PCPs. In addition, some medical groups use hospitalists to avoid the rising malpractice cost of caring for sick, complex inpatients.[6]

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20069

© 2008 Mount Sinai School of Medicine

There is anecdotal evidence that hospitals are using hospitalists, physician assistants, and nurse practitioners to accommodate changes in resident duty hours. Hospitalists have also allowed hospitals to comply with Joint Commission on Accreditation of Healthcare Organizations safety standards.[7]

COST EFFECTIVENESS AND CLINICAL OUTCOMES

Cost reduction was the major driving force for the hospitalist movement, and most of the medical literature has focused on cost reduction by hospitalists (Table 1).

The majority of the studies have been observational trials. There have been 3 prospective quasi-experimental trials.[8–10] All the studies compared hospitalists to either faculty physicians who attended on the wards an average of 1 month per year or community physicians who divided their time between inpatient and outpatient work. Almost every study found lower costs and shorter length of stay in patients cared for by hospitalists with no differences in clinical outcomes such as mortality and readmission rates.

The literature is rich in observational studies but is lacking in randomized controlled trials. Most trials have studied outcomes for a single hospital; however, a recent article in the *New England Journal of Medicine* reported outcomes from 45 hospitals nationwide and found that hospitalists achieved lower costs and shorter length of stay with no difference in clinical outcomes.[11]

There have also been studies looking at the cost effectiveness and clinical outcomes of hospitalist care for specific disease entities such as community-acquired pneumonia and congestive heart failure. In 2 studies of patients with community-acquired pneumonia, hospitalists achieved shorter lengths of stay and lower costs with no difference in clinical outcomes.[12,13] Rifkin *et al.*[12] found that hospitalists called fewer infectious disease consultations. In 1 study, patients with community-acquired pneumonia admitted to a critical care hospitalist team had higher costs than patients cared for by PCPs.[14]

Several studies have found that hospitalists achieve lower costs and shorter lengths of stay for patients admitted with congestive heart failure.[15,16] Lindenauer *et al.*[15] found that hospitalists were more likely to prescribe an angiotensin converting enzyme inhibitor or angiotensin receptor blocker within 24 hours of admission. Roytmann *et al.*[16] found that hospitalists ordered fewer serial radiographs and consultations.

QUALITY OF THE LITERATURE

The majority of studies looking at the cost effectiveness and quality of hospitalist services have been observational studies within a single hospital or a small group of hospitals. There have been no randomized controlled trials. However, hospital-based medicine has quickly become an accepted model of care, and the peer-reviewed literature has not been able to keep up with the changes in the field.

Analysis models from the manufacturing industry may be an effective way of determining the cost and quality of the hospitalist model. Six Sigma is a model used to analyze and improve manufacturing and flow in many companies and has been used in our hospital to improve the quality of medical care. Six Sigma uses a specified methodology to analyze data and implement changes to improve process flow and decrease errors. Six Sigma may be a good model for hospitals to use when they are further analyzing the quality of their hospitalists services.[17]

One area that has not been addressed is the possibility of cost shifting from the inpatient setting to the outpatient setting. Longitudinal or large-scale database analyses may help clarify whether the cost reductions of hospitalists are maintained after discharge in the outpatient setting.

CONTINUITY OF CARE

One of the concerns with the growth of hospitalists is the potential for problems with continuity of care and lack of effective communications with PCPs. The Joint Commission on Accreditation of Healthcare Organizations has set communication as a 2008 patient safety goal and recommends a standard approach to handoff communications.[18]

There have been a number of studies looking at the communication between hospitalists and PCPs during the transition from the inpatient setting to the outpatient setting. In a systematic review of the literature, Kripalani *et al.*[19] found that direct communication between hospitalists and PCPs was rare (3%–20%) and that the availability of the discharge summary at the time of the first outpatient visit was low (12%–34%). They also found that discharge summaries lacked relevant test results, discharge medications, family counseling, pending test results, and follow-up plans.

Goldman *et al.*[20] proposed that hospitalists should follow 6 rules for discharge communication. Hospitalists should provide accurate and timely communication of diagnostic findings, treatment, complications, consultations, pending tests, and follow-up plans to PCPs.

Table 1. *Cost, Length of Stay, and Clinical Comparison of Hospitalists Versus Nonhospitalists.*

Study (Year)	Setting	Design	Comparison Group(s)	Sample Size	Hospital Costs	Risk-Adjusted Length Of Stay	Outcome
Wachter *et al.*[8] (1998)	Teaching hospital	Alternate-day controlled trial	Traditional teaching service	1623 patients	Lower by $700-1071	Shorter by 0.6 days	No difference
Diamond *et al.*[24] (1998)	Community-based teaching hospital	Retrospective cohort	PCPs	9935 patients	Lower by $587	Shorter by 1 day	Reduction in 14-day readmission
Craig *et al.*[25] (1999)	16 HMO hospitals	Retrospective cohort	PCPs	Not available	Year 1: Higher by 13% Year 2: Higher by 5%	11% lower	Not reported
Davis *et al.*[26] (2000)	Rural community hospital	Retrospective cohort	Internists	2124 patients	Lower by $560	Shorter by 1.4 days	No difference
Halpert *et al.*[27] (2000)	Urban teaching hospital	Retrospective cohort	Faculty physicians	2265 patients	Lower by $426	Shorter by 0.5 days	No difference
Hackner *et al.*[28] (2001)	Community-based teaching hospital	Retrospective cohort	General internists	1637 patients	Lower by $851	Shorter by 1 day	No difference
Kearns *et al.*[9] (2001)	Suburban county hospital	Prospective quasi-experimental	Clinic-based internists	4456 patients	No difference	No difference	No difference
Palmer *et al.*[10] (2001)	Academic teaching hospital	Prospective quasi-experimental	Faculty physicians: general internists and specialists	2464 patients	Lower by $1777	Shorter by 0.8 days	Mortality higher in specialist group
Tingle and Lambert[29] (2001)	Community-based teaching hospital	Retrospective cohort	FMTA	652 patients	No difference	No difference	No difference
Molinari and Short[30] (2001)	HMO hospital	Retrospective cohort	Network PCPs	1,319 patients	Not studied	Fewer admissions with unnecessary days	Not studied
Auerbach *et al.*[31] (2002)	Community-based teaching hospital	Retrospective cohort	Community physicians	5,308 patients	Year 1: no difference Year 2: lower by $822	Year 1: no difference Year 2: shorter by 0.61 days	Lower in-hospital mortality
Kaboli *et al.*[32] (2004)	Academic teaching hospital	Prospective, quasi-experimental	Faculty physicians	1,706 admissions	10% reduction	Shorter by 1 day	No difference
Everett *et al.*[33] (2004)	Community hospital	Retrospective cohort	General internists	11,750 admissions	8.3% reduction	16.1% reduction	No difference

Rifkin *et al.*[34] (2004)	Academic teaching hospital	Retrospective cohort	Faculty physicians	11,388 admissions	Not studied	49% less likely to exceed the average length of stay for each diagnosis-related group	Not studied
Halasyamani *et al.*[35] (2005)	Community-based teaching hospital	Retrospective cohort	Community physicians, private hospitalists, and academic hospitalists	10,595 patients	10% reduction with academic hospitalists and 6% reduction with private hospitalists versus community physicians	20% reduction with academic hospitalists and 8% reduction with private hospitalists versus community physicians	No difference
Lindenauer *et al.*[11] (2007)	45 hospitals nationwide	Retrospective cohort	General internists and family physicians	76,926 patients	Lower by $268 versus general internists and by $125 versus family physicians	Shorter by 0.4 days versus both groups	No difference
Southern *et al.*[36] (2007)	Academic teaching hospital	Retrospective cohort	Faculty physicians	9,037 discharges	Not studied	Shorter length of stay by 0.8 days (reduction greatest for patients requiring close clinical monitoring)	No difference
Everett *et al.*[37] (2007)	Community-based teaching hospital	Retrospective cohort	Community physicians, private hospitalists, and academic hospitalists	57,174 admissions	30% reduction with academic hospitalists and 24% reduction with private hospitalists versus community physicians	40% reduction with academic hospitalists and 30% reduction with private hospitalists versus community physicians	2.3–2.6% greater 30-day admission with academic hospitalists versus community physicians
Carek *et al.*[38] (2008)	For-profit community hospital	Retrospective cohort	FMTA and community physicians	5,453 admissions	Lower for FMTA	Shorter for FMTA	No difference

Abbreviations: FMTA, family medicine teaching attending; HMO, health maintenance organization; PCP, primary care physician.

HOSPITALIST WORKFORCE

The growth in the hospitalist movement has left questions about the sustainability of the workforce. Three-quarters of hospitalists are general internists.[21] However, according to the American College of Physicians, primary care is facing a shortage of physicians, particularly in light of the aging baby boomer population.[22] Although the hospitalist workforce initially benefited from the movement of established PCPs to hospitalist roles, it will likely face similar workforce shortages. In fact, a 2007 survey of medical school graduates found that only 5.1% planned to practice general internal medicine.[21] Burnout may also be an issue. Wachter and Goldman[1] suggested that 3 to 6 months of inpatient service per year is more sustainable than 12 months of full-time hospitalist service.

Financial incentives for hospitalists may improve the sustainability of the hospitalist workforce. Several for-profit hospitalist staffing companies have had lucrative profits that may be passed on to their hospitalists. IPC, the largest hospitalist staffing company, became a publicly traded corporation in early 2008 with a market cap of $330 million and annual revenues of $190 million in 2007. However, IPC's portion of the hospitalist market is very small.[23]

CONCLUSION

It is clear that the hospitalist movement has already transformed inpatient medicine. The literature is clear that hospitalists lower costs and shorten the length of hospital admissions without diminishing the quality of patient care. Future challenges for the field lie in improving transitions of care and communication with PCPs, enhancing patient quality, and ensuring the sustainability of the workforce.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Wachter RM, Goldman L. The emerging role of "hospitalists" in the American health care system. *N Engl J Med* 1996; 335: 514–517.
2. Auerbach AD, Nelson EA, Lindenauer PK, et al. Physician attitudes toward and prevalence of the hospitalist model of care: results of a national survey. *Am J Med* 2000; 109: 648–653.
3. Kralovec PD, Miller JA, Wellikson L, Huddleston JM. The status of hospital medicine groups in the United States. *J Hosp Med* 2006; 1: 75–80.
4. Lurie JD, Miller DP, Lindenauer PK, et al. The potential size of the hospitalist workforce in the United States. *Am J Med* 1999; 106: 441–445.
5. Wachter RM, Goldman L. The hospitalist movement 5 years later. *JAMA* 2002; 287: 487–494.
6. Pham HH, Devers KJ, Kuo S, et al. Health care market trends and the evolution of hospitalist use and roles. *J Gen Intern Med* 2005; 20: 101–107.
7. Dunn A (Mt. Sinai Hospital, New York, NY). Personal communication, May, 2008.
8. Wachter RM, Katz P, Showstack J, et al. Reorganizing an academic medical service: impact on cost, quality, patient satisfaction, and education. *JAMA* 1998; 279: 1560–1565.
9. Kearns PJ, Wang CC, Morris WJ, et al. Hospital care by hospital-based and clinic-based faculty: a prospective, controlled trial. *Arch Intern Med* 2001; 161: 235–241.
10. Palmer HC, Armistead NS, Elnicki DM, et al. The effect of a hospitalist service with nurse discharge planner on patient care in an academic teaching hospital. *Am J Med* 2001; 111: 627–632.
11. Lindenauer PK, Rothberg MB, Pekow PS. Outcomes of care by hospitalists, general internists, and family physicians. *N Engl J Med* 2007; 357: 2589–2600.
12. Rifkin WD, Conner D, Silver A, Eichorn A. Comparison of processes and outcomes of pneumonia care between hospitalists and community-based primary care physicians. *Mayo Clin Proc* 2002; 77: 1053–1058.
13. Stein MD, Hanson S, Tammaro D, et al. Economic effects of community versus hospital-based faculty pneumonia care. *J Gen Intern Med* 1998; 13: 774–777.
14. Smith PC, Westfall JM, Nichols RA. Primary care family physicians and 2 hospitalist models: comparison of outcomes, processes, and costs. *J Fam Pract* 2002; 51: 1021–1027.
15. Lindenauer PK, Chehabeddine R, Pekow P, et al. Quality of care for patients hospitalized with heart failure: assessing the impact of hospitalists. *Arch Intern Med* 2002; 162: 1251–1256.
16. Roytman MM, Thomas SM, Jiang CS. Comparison of practice patterns of hospitalists and community physicians in the care of patients with congestive heart failure. *J Hosp Med* 2008; 3: 35–41.
17. Chassin R. The Six Sigma Initiative at Mount Sinai Medical Center. *Mt Sinai J Med* 2008; 75: 45–52.
18. Joint Commission on Accreditation of Healthcare Organizations. Joint Commission National patient safety goals. Available at: http://www.jointcommission.org/patientsafety/nationalpatientsafetygoals/08_hap_npsgs.htm. Accessed April 2008.
19. Kripalani S, LeFevre F, Phillips CO, et al. Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. *JAMA* 2007; 297: 831–841.
20. Goldman L, Pantilat SZ, Whitcomb WF. Passing the clinical baton: 6 principles to guide the hospitalist. *Am J Med* 2001; 111: 36S–39S.
21. McMahon LF Jr. The hospitalist movement–time to move on. *N Engl J Med* 2007; 357: 2627–2629.
22. American College of Physicians. Addressing healthcare workforce issues for the future. Available at:

http://www.acponline.org/advocacy/where_we_stand/workforce/pc_workforce.pdf. Accessed April 2008.

23. Goldstein, J. Riding the hospitalist boom to an IPO. Available at: http://blogs.wsj.com/health/2008/03/31/riding-the-hospitalist-boom-to-an-ipo/?mod=wsjblog. Accessed April 2008.
24. Diamond HS, Goldberg E, Janosky JE. The effect of full-time faculty hospitalists on the efficiency of care at a community teaching hospital. *Ann Intern Med* 1998; 129: 197–203.
25. Craig DE, Hartka L, Likosky WH, et al. Implementation of a hospitalist system in a large health maintenance organization: the Kaiser Permanente experience. *Ann Intern Med* 1999; 130: 355–359.
26. Davis KM, Koch KE, Harvey JK, et al. Effects of hospitalists on cost, outcomes, and patient satisfaction in a rural health system. *Am J Med* 2000; 108: 621–626.
27. Halpert AP, Pearson SD, LeWine HE, et al. The impact of an inpatient physician program on quality, utilization, and satisfaction. *Am J Manag Care* 2000; 6: 549–555.
28. Hackner D, Tu G, Braunstein GD, et al. The value of a hospitalist service: efficient care for the aging population? *Chest* 2001; 119: 580–589.
29. Tingle LE, Lambert CT. Comparison of a family practice teaching service and a hospitalist model: costs, charges, length of stay, and mortality. *Fam Med* 2001; 33: 511–515.
30. Molinari C, Short R. Effects of an HMO hospitalist program on inpatient utilization. *Am J Manag Care* 2001; 7: 1051–1057.
31. Auerbach AD, Wachter RM, Katz P, et al. Implementation of a voluntary hospitalist service at a community teaching hospital: improved clinical efficiency and patient outcomes. *Ann Intern Med* 2002; 137: 859–865.
32. Kaboli PJ, Barnett MJ, Rosenthal GE. Associations with reduced length of stay and costs on an academic hospitalist service. *Am J Manag Care* 2004; 10: 561–568.
33. Everett GD, Anton MP, Jackson BK, et al. Comparison of hospital costs and length of stay associated with general internists and hospitalist physicians at a community hospital. *Am J Manag Care* 2004; 10: 626–630.
34. Rifkin WD, Holmboe E, Scherer H, et al. Comparison of hospitalists and nonhospitalists in inpatient length of stay adjusting for patient and physician characteristics. *J Gen Intern Med* 2004; 19: 1127–1132.
35. Halasyamani LK, Valenstein PN, Friedlander MP, et al. A comparison of two hospitalist models with traditional care in a community teaching hospital. *Am J Med* 2005; 118: 536–543.
36. Southern WN, Berger MA, Bellin EY, et al. Hospitalist care and length of stay in patients requiring complex discharge planning and close clinical monitoring. *Arch Intern Med* 2007; 167: 1869–1874.
37. Everett G, Uddin N, Rudloff B. Comparison of hospital costs and length of stay for community internists, hospitalists, and academicians. *J Gen Intern Med* 2007; 22: 662–667.
38. Carek PJ, Boggan H, Mainous AG, et al. Inpatient care in a community hospital: comparing length of stay and costs among teaching, hospitalist, and community services. *Fam Med* 2008; 40: 119–124.

Hospitalist Educators: Future of Inpatient Internal Medicine Training

Jill Goldenberg, MD[1], and Jeffrey J. Glasheen, MD[2]

[1]Division of General Internal Medicine, Department of Medicine, Mount Sinai School of Medicine, New York, NY
[2]Hospital Medicine Section, Division of General Internal Medicine, Department of Medicine, University of Colorado at Denver Health Sciences Center, Denver, CO

ABSTRACT

Academic hospitalists have grown in number and influence over the past decade. This has fueled concerns about the effect of hospitalists on resident and student education. While the bulk of the literature favors the hospitalist teaching model to a more traditional model concerns remain that hospitalists may negatively imipact housestaff autonomy and reduce exposure to subspeciality physicians. This paper will review the literature exploring the effect of the hospitalist teaching model on resident and student education. ***Mt Sinai J Med 75:430–435, 2008.*** © *2008 Mount Sinai School of Medicine*

Hospital medicine is the fastest growing medical specialty in the American healthcare system. Its providers currently number approximately 20,000, with some estimates calling for twice that number when the field matures.[1,2] However, unlike the majority of clinical advancements, the hospital medicine movement developed more prominently in the community sector and was translated back to academics secondarily. This earlier adoption in the community can be attributed to the economic drivers of the field, namely managed care pressures and the need to staff unassigned patients admitted through the Emergency Department.[3,4] Teaching hospitals were more sheltered from these economic drivers with their funded teaching mission and housestaff availability to care for unassigned Emergency Department patients.

Another equally important reason for the slow adoption of academic hospitalists is the concern over hospitalists' impact on medical student and resident education. Educators initially voiced apprehension over the unintended consequences of replacing seasoned educators with young, untested hospitalists. These anxieties were heightened by the perception that these more "community-minded" physicians might not be qualified to act as housestaff and student educators and role models. Further concerns about the reduction in resident autonomy and satisfaction with full-time hospitalists were expressed. Others opined that hospitalists, with a mandate to focus on efficiency, might reduce length of stay and resource utilization, and this could negatively impact trainees' ability to learn from their patients' illnesses. Additionally, hospitalists' comfort level in managing inpatient medical problems might result in a reduction in the number of consultations, thereby tainting the fellowship trainees' experience. Finally, subspecialists worried about the effect of their reduced attending exposure on subspecialty recruitment and the development of medical researchers.[4] In this article, we investigate the impact of the hospitalist model on the educational experiences of medical and pediatric resident and student learners.

Address Correspondence to:

Jill Goldenberg
Division of General Internal Medicine
Department of Medicine
Mount Sinai School of Medicine
New York, NY
Email: jill.goldenberg@mountsinai.org

RESIDENT OUTCOMES

The first published experience with academic internal medicine (IM) hospitalists noted significant improvements in length of stay and cost of care with no differences in mortality or readmission rates at the University of California, San Francisco

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20075

© 2008 Mount Sinai School of Medicine

(UCSF). In terms of educational outcomes, the authors found a trend toward increased resident satisfaction with hospitalist attendings compared to nonhospitalist attendings (60% versus 30% were "very satisfied," $P = 0.06$). Furthermore, 68% and 13% of the housestaff on the hospitalist and nonhospitalist services, respectively, reported that cost-effective medicine "was emphasized very much" ($P < 0.001$) without differences in other educational parameters such as the instruction of pathophysiology or evidence-based medicine. The authors concluded that the hospitalist teaching model improved resource utilization without negatively effecting clinical or educational outcomes[3] (Table 1).

Similar findings were revealed when Brigham and Women's Hospital adopted an IM hospitalist teaching model. In the preintervention/postintervention evaluation, length of stay decreased significantly (from 3.5 to 3.0 days, $P < 0.001$) along with costs ($426 less per patient), whereas resident satisfaction with the overall educational experience increased from 3.20 to 3.44 on a 4-point Likert scale ($P = 0.04$). The authors noted that informal feedback suggested that the higher satisfaction stemmed from a "more cohesive team structure and real-time, practical teaching fostered by the presence of a hospital-based attending."[5,6]

At the University of Chicago, 1 of 4 teaching IM inpatient services was transformed into a hospitalist-run service staffed by 2 to 4 full-time internists who ran the service individually 3 to 6 months of the year. This model was compared to the 3 traditional services. Eighty-six housestaff members rotating on the 2 models were surveyed with a 5-point Likert scale. In comparison with nonhospitalist attendings, residents felt that hospitalist attendings provided a better learning experience ($P = 0.02$), higher quality attending rounds ($P = 0.02$), and a higher emphasis on education ($P < 0.01$). In an effort to better understand housestaff perceptions of the hospitalist model (versus the individual attendings), the authors compared the survey results of residents who rotated with a hospitalist versus those who did not. They found that residents who rotated with a hospitalist were less likely to believe that their autonomy was limited (8% versus 28%, $P = 0.02$) or that their exposure to diverse faculty teaching styles and research interests was compromised (41% versus 60%, $P = 0.08$) by working with hospitalist attendings.[7]

The positive effect of hospitalists as educators appears to extend to community teaching hospitals as well. Kulaga *et al.*[8] investigated the impact of hospitalist clinician-educators on a community-based IM training program by comparing 2 hospitalist clinician-educators to practicing private physicians who provided both formal and informal bedside teaching. They reported that hospitalist length of stay and cost per case were reduced 20.8% and 18.4%, respectively, versus the traditional model. In terms of educational outcomes, 36 residents were surveyed with a 5-point Likert scale. Of the 67 out of 86 housestaff members who responded, 97% reported that the hospitalist educators "improved" or "greatly improved" teaching in several settings, including bedside rounds, attending rounds, and didactics. Furthermore, 90% of respondents noted that hospitalists encouraged the use of evidence-based medicine. Interestingly, despite the significant cost savings per patient on the hospitalist service, very few residents noted a difference in the number of tests ordered when they were working with the hospitalist attendings. In terms of autonomy, while working with hospitalist staff, 75% of the housestaff felt that their roles as team leaders either stayed unchanged or increased. Overall, 100% of respondents preferred the new hospitalist-run model.

Similar findings have been reported from pediatric hospitalist systems. Landrigan *et al.*[9] surveyed pediatric interns and residents before and after the introduction of a hospitalist system at a free-standing academic children's hospital. The interns' overall rating of the ward rotation increased from 4.1 to 4.7 on a 5-point Likert scale ($P = 0.01$) after the introduction of the program, whereas their ratings of comparison rotations did not change over the same time. Additional benefit was seen in their level of satisfaction with the educational experience, their quality of life, and their self-assessment of their skills and knowledge of general pediatrics. The hospitalist attendings were rated more highly than nonhospitalist attendings as role models and teachers and were seen as more knowledgeable, accessible, and involved in learning and feedback processes.

Another innovative pediatric model, in which hospitalists triaged patients to the hospitalist service or the teaching service according to the educational value of the cases, was associated with higher resident ratings and satisfaction. When complex patients were admitted after hours, hospitalists were available and responsible for bedside teaching and guidance. Overall, the average resident satisfaction with the inpatient rotation increased from 4.4 to 5.6 on a 6-point scale after the introduction of this program.[10]

Hauer *et al.*[11] were the first to evaluate the effect of IM hospitalists at more than 1 hospital, albeit in the same academic system. A total of 1587 evaluations by medical students and residents of 17 hospitalists and 52 traditional attendings

Table 1. *Outcome Studies of the Hospitalist Educator Model.*

Learners, Institution, and Source	Intervention	Outcomes
• Residents • Moffitt-Long Hospital at UCSF, San Francisco, CA • Wachter *et al.*[3] (1998)	Hospitalist faculty versus nonhospitalist faculty	• 55 surveys completed, 56% response rate • More satisfaction with hospitalist attendings than nonhospitalist attendings† • Hospitalists emphasized cost-effective medicine.* • No differences in other educational parameters
• Residents • Brigham and Women's Hospital, Boston, MA • Halpert *et al.*[5] (2000)	Precomparison and postcomparison of teaching attendings without direct patient care responsibilities versus hospitalists	• 49 preintervention and 240 postintervention surveys rose from 4.4 to 5.6 (6-point scale).** • Increase in overall educational experience with hospitalists*
• Residents • Valley Children's Hospital, Fresno, CA • Wilson[10] (2001)	Hospitalists triaged patients to the teaching or hospitalist service on the basis of "educational value."	• Average evaluation scores rose from 4.4 to 5.6 (6-point scale). • Better educational case mix
• Residents • University of Chicago Pritzer School of Medicine, Chicago, IL • Chung *et al.*[7] (2002)	One of 4 teaching inpatient services was transformed into a hospitalist-run service and compared to 3 nonhospitalist teaching services	• 86 surveyed, 78% response rate • Hospitalist attendings provided a better learning experience,* higher quality attending rounds,* and a higher emphasis on education* • Residents on the hospitalist service were less likely to believe their autonomy* or exposure to diverse faculty and teaching styles† was limited
• Residents • Children's Hospital, Boston, MA • Landrigan *et al.*[9] (2002)	Comparison of teaching before and after the introduction of a hospitalist system	• Surveys completed by 67 interns and 54 residents, 61% response rate • Intern ratings of the ward rotation,* quality of life,* and self assessed knowledge* increased • Hospitalist attendings were more highly rated as role models and teachers* and were seen as more knowledgeable,* accessible,* and involved in learning and feedback processes.* • Bedside teaching was rated lower for hospitalist attendings*
• Residents and students • 2 UCSF hospitals, San Francisco, CA • Hauer *et al.*[11] (2004)	Comparison of 17 hospitalists and 52 traditional attendings	• 1587 surveys completed, 91% response rate • Hospitalists provided higher overall satisfaction, teaching effectiveness, and perception of educational value* • No significant differences in interest in teaching, availability, or focus on cost effectiveness

(*continued overleaf*)

Table 1. *(Continued)*.

Learners, Institution, and Source	Intervention	Outcomes
• Residents and students • Grady Memorial Hospital, Atlanta, GA • Kripalani *et al.*[12] (2004)	Comparison of 12 medical ward teams, 6 staffed by hospitalists, 3 by traditional general internists, and 3 by subspecialists	• 342 surveys completed, 81% response rate • Both hospitalists and general internist attendings yielded higher quality teaching rounds, taught more enthusiastically, used evidence-based medicine, role-modeled good bedside manner, and had a greater presence and involvement with patient care than subspecialist attendings. • Hospitalist mean scores were higher than those of subspecialists*
• Residents • Norwalk Hospital, Norwalk, CT • Kulaga *et al.*[8] (2004)	Comparison of 2 hospitalist clinician-educators to practicing private physicians	• 28 surveys completed, 78% response rate • 97% of residents reported that hospitalist educators "improved" or "greatly improved" teaching • 90% noted that hospitalists encouraged their use of evidence-based medicine • 75% of housestaff felt that the roles as team leaders either stayed unchanged or increased; 25% experienced decreased autonomy • 100% of respondents preferred the new hospitalist-run model
• Students • Oregon Health and Science University School of Medicine, Portland, OR • Hunter *et al.*[13] (2004)	Hospitalist teaching service versus nonhospitalist-staffed teaching service	• 99 surveys completed, 72% response rate • Hospitalists were rated higher in communication of goals, teaching effectiveness, and combined parameters*
• Students • Penn State Children's Hospital and Hershey Medical Center, Hershey, PA • Geskey and Kees-Folts[14] (2007)	Hospitalist and nonhospitalist faculty	• 67 surveys completed, 100% response rate • Hospitalists were rated as more effective teachers, better pediatricians, and more effective student advocates and were given higher overall evaluation scores*

Abbreviation: UCSF, University of California, San Francisco.

* Statistically significant finding in favor of hospitalists. † Nonstatistically significant finding in favor of hospitalists.

** Statistics not provided in the original paper.

over a 2-year period at 2 UCSF-affiliated hospitals showed that hospitalists were rated higher as role models, were felt to be more effective teachers, and were perceived as having more knowledge of inpatient medicine and pathophysiology. Overall, resident evaluators reported higher satisfaction with hospitalist attendings (8.3 versus 8.0 on a 9-point Likert scale, $P < 0.001$) and higher perception of the educational value of the rotation (3.9 versus 3.7 on a 5-point Likert scale, $P = 0.04$). Although these findings were statistically significant, the absolute difference in endpoints was small, resulting in indefinite educational significance. Nonetheless, in this large study, it was clear that the hospitalist model was at least as effective as the traditional model.

In the only study to specifically evaluate hospitalists versus general internists and subspecialty medicine attendings, Kripalani *et al.*[12] surveyed a total of 423 medical students and residents at an Emory-affiliated hospital. Of the 12 medical ward teams, 6 were staffed by hospitalists, 3 were staffed by traditional general internists, and 3 were staffed by subspecialists. In all, 63 attendings were evaluated. Hospitalists and general internists were rated higher then subspecialists and were considered more effective teachers ($P = 0.03$ and 0.01, respectively). The qualitative evaluations were more positive for hospitalists than for general medicine and subspecialty faculty. Specifically, the evaluators remarked that IM hospitalists provided quality teaching rounds, taught enthusiastically, used evidence-based medicine, role-modeled good bedside manner, and had a greater presence and involvement with patient care.

MEDICAL STUDENT OUTCOMES

In addition to the articles by Hauer *et al.*[11] and Kripalani *et al.*,[12] 2 additional studies have directly examined medical student satisfaction with the hospitalist attending teaching model. Hunter *et al.*[13] surveyed 138 medical students at a university hospital in Oregon and found that hospitalists were rated as more effective teachers than traditional faculty. Students felt that hospitalists were better able to communicate clear goals for the rotation ($P = 0.011$), were more effective as clinical teachers ($P = 0.016$), and were more enthusiastic educators than nonhospitalist attendings. Furthermore, the combination of all parameters in the evaluation favored hospitalists over nonhospitalists ($P < 0.001$).

A more recent article[14] evaluated third-year medical student perceptions of hospitalist and nonhospitalist faculty during a pediatric inpatient clerkship and found that hospitalists provided more effective teaching and an overall better experience. Specifically, the hospitalists were rated as more effective teachers than nonhospitalists (3.87 versus 2.91 on a 4-point scale; $P < 0.001$). Furthermore, hospitalists were rated as better pediatricians (3.94 versus 3.25, $P < 0.001$) and more effective student advocates (3.76 versus 2.97, $P < 0.001$) and were given a higher overall evaluation (3.93 versus 3.06, $P < 0.001$).

NEGATIVE CONSEQUENCES OF THE HOSPITALIST TEACHING MODEL

The bulk of the hospitalist educational outcomes to date have been positive. Still, it is important to note that the study designs have generally favored single sites and evaluated subjective self-perception survey data rather than objective outcomes. As such, we have a limited understanding of the implications of hospitalist teachers on hard educational endpoints such as board scores, skills acquisition, and patient care outcomes.

Moreover, studies have uncovered negative consequences of the hospitalist teaching model. For example, Landrigan *et al.*'s study[9] of pediatric housestaff revealed 3 potentially negative outcomes. First, the hospitalist attendings' bedside teaching skills were rated lower by all housestaff (3.7 versus 4.2, $P < 0.001$). Second, the senior residents reported very little positive effect after the change to the hospitalist model. Third, the senior residents reported a statistically nonsignificant trend toward a lower level of autonomy in clinical decision making after the introduction of the new model. Specifically, ratings related to independent decision making decreased from 4.1 to 3.5 ($P = 0.7$), and ratings of opportunity to learn supervisory skills also decreased from 4.2 to 3.7 ($P = 0.7$). Lack of autonomy was also noted by Kulaga *et al.*,[8] who found that 25% of residents felt that their autonomy was decreased by the introduction of a hospitalist model. Landrigan and colleagues concluded that these changes in perceived autonomy and supervision likely resulted from an increased level of communication and collaboration with the hospitalist attendings that was not present in the previous traditional system. It is unclear how to address these perceived decreases in autonomy. In many cases, this increase in supervision and attending involvement is a byproduct of the changing landscape of healthcare and likely represents a positive move toward improved patient safety and quality outcomes. That being said, future investigation into the impact of hospitalist attendings should evaluate methods to maintain resident decision making in a manner that promotes active learning.

Another potential negative trend of the growth in the hospitalist teaching model is a move away from subspecialist involvement in medical residency training. Although an unproven concern, it is reasonable to suspect that a decrease in subspecialty attending exposure on the general medical services, especially at the student level, could negatively impact academic subspecialty and basic science research recruitment. In response to this potential reduction in subspecialty contact, UCSF has instituted a multitiered program aimed at reengaging subspecialty faculty in housestaff training. This program includes dedicated subspecialty morning reports, a rotating subspecialty journal club, fellowship events, an intern-year subspecialty elective, and a monthly "bench to bedside" conference featuring biomedical researchers.[15]

SUMMARY

The reach of academic hospitalists has grown rapidly over the past decade, and this has fueled concerns about the influence of hospitalists on resident and student education. The literature to date suggests that hospitalist educators are rated favorably by both housestaff and medical students. Indeed, in most studies, hospitalists are rated more positively than their corresponding subspecialist and nonhospitalist general internist and pediatric attendings. Hospitalists appear to serve as excellent role models, providing high-quality teaching, a focus on evidence-based medicine, and ready availability. However, appropriate concerns about resident autonomy and the impact on the role that subspecialists play in medical education exist. As we forge into the 21st century, it is clear that hospitalists will play a large and likely increasing role in residency and student education. Although the known literature would suggest that this is a positive trend, it is incumbent upon medical educators to avoid unintended consequences and ensure that the benefits of the hospitalist model persist.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Society of Hospital Medicine. Growth of hospital medicine nationwide. Available at: http://www.hospitalmedicine.org/content/navigation-menu/media/growthofhospitalmedicinenationwide/growth_of_hospital_m.htm. Accessed March 2008.
2. Wachter RM. Reflections: the hospitalist movement a decade later. *J Hosp Med* 2006; 1: 248–252.
3. Wachter RM, Katz P, Showstack J, et al. Reorganizing an academic medical service: impact on cost, quality, patient satisfaction, and education. *JAMA* 1998; 279: 1560–1565.
4. Goldman L. The impact of hospitalists on medical education and the academic health system. *Ann Intern Med* 1999; 130(pt 2): 364–367.
5. Halpert AP, Pearson SD, LeWine HE, Mckean SC. The impact of an inpatient physician program on quality, utilization, and satisfaction. *Am J Manag Care* 2000; 6: 549–555.
6. Brown MD, Halpert A, McKean S, et al. Assessing the value of hospitalists to academic health centers: Brigham and Women's Hospital and Harvard Medical School. *Am J Med* 1999; 106: 134–137.
7. Chung P, Morrison J, Jin L, et al. Resident satisfaction on an academic hospitalist service: time to teach. *Am J Med* 2002; 112: 597–601.
8. Kulaga ME, Charney P, O'Mahony SP, et al. The positive impact of initiation of hospitalist clinician educators. *J Gen Intern Med* 2004; 19: 293–301.
9. Landrigan CP, Muret-Wagstaff S, Chiang VW, et al. Effect of a pediatric hospitalist system on housestaff education and experience. *Arch Pediatr Adolesc Med* 2002; 156: 877–883.
10. Wilson SD. Employing hospitalists to improve residents' inpatient learning. *Acad Med* 2001; 76: 556.
11. Hauer KE, Wachter RM, McCulloch CE, et al. Effects of hospitalist attending physicians on trainee satisfaction with teaching and with internal medicine rotations. *Arch Intern Med* 2004; 164: 1866–1871.
12. Kripalani S, Pope AC, Rask K, et al. Hospitalists as teachers. How do they compare to subspecialty and general medicine faculty? *J Gen Intern Med* 2004; 19: 8–15.
13. Hunter AJ, Desai SS, Harrison RA, Chan BK. Medical student evaluation of the quality of hospitalist and nonhospitalist teaching faculty on inpatient medicine rotations. *Acad Med* 2004; 79: 78–82.
14. Geskey JM, Kees-Folts D. Third-year medical students' evaluation of hospitalist and nonhospitalist faculty during the inpatient portion of their pediatrics clerkships. *J Hosp Med* 2007; 2: 17–22.
15. Hollander H. Response to the effect of hospitalist systems on residency education: re-incorporating medical subspecialists. *Acad Med* 2001; 76: 555–556.

Achieving Hospital Medicine's Promise Through Internal Medicine Residency Redesign

Jeffrey J. Glasheen, MD[1], Jill Goldenberg, MD[2], and John R. Nelson, MD[3]

[1]Hospital Medicine Section, Division of General Internal Medicine, Department of Medicine, University of Colorado at Denver Health Sciences Center, Denver, CO
[2]Division of General Internal Medicine, Department of Medicine, Mount Sinai School of Medicine, New York, NY
[3]Overlake Hospital Medical Center, Bellevue, WA

ABSTRACT

The promise of the hospital medicine movement is that the hospitalist model of care will provide better outcomes than the system it replaced. This means improving the quality and processes of care, reducing inefficiencies and lowering costs. Despite some documented improvements in these areas hospitalists have yet to achieve their pinnacle. These shortfalls likely result from training providers in residencies that have yet to evolve to address the specific needs of hospitalists. While most internal medicine residency training programs stress inpatient care they underemphasize key components of a successful hospitalist career. This paper overviews the state of the hospitalist movement, the current educational training deficiencies and the methods to deliver hospitalist-focused training. ***Mt Sinai J Med 75:436–441, 2008.*** © *2008 Mount Sinai School of Medicine*

Key Words: hospitalist, hospital medicine, medical education, hospitalist education, hospitalist training, internal medicine, internal medicine training, residency redesign.

Address Correspondence to:

Jeffrey J. Glasheen
Division of General Internal Medicine
Department of Medicine
University of Colorado at Denver Health Sciences Center
Denver, CO
Email: jeffrey.glasheen@ucdenver.edu

The promise of the hospital medicine movement is that the hospitalist model will improve outcomes in comparison with the traditional model of care, in which a primary care doctor or medical subspecialist cares for his outpatients admitted to the hospital. This means reducing costs, promoting efficiencies, enhancing patient safety, and improving the overall quality of hospital care.

To date, minimal effort has gone into training internal medicine (IM) residents to achieve the outcomes required of hospitalists. This likely stems from the belief that IM training, historically focused principally on inpatient care, provides ideal training for future hospitalists. For the field to achieve its potential, we must alter the training landscape to recognize the unique attributes of the burgeoning new specialty of hospital medicine. This article develops the case for specialized training in hospital medicine, outlines the recommended components of this training, and reviews the mechanisms for delivering this content. Although we focus on IM training, many of the concepts are applicable to other fields of medicine in which hospitalists are trained, such as family and pediatric medicine.

HOSPITALIST OUTCOMES

The data supporting the hospitalist model are generally positive. However, they are temporally disparate in the size of the effect. Early data revealed an average decreased length of stay of 17% with a 13% reduction in hospital costs per case[1] and reductions in short-term mortality rates.[2,3] Furthermore, a hospitalist-orthopedic comanagement model was associated with significantly fewer minor

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20077

© 2008 Mount Sinai School of Medicine

complications.[4] More recent data, however, have shown less robust effects, with 1 study revealing no improvement in outcomes when hospitalist consultants were involved in the care of surgical patients and another large hospitalist study showing more modest 12% and 5% reductions in length of stay and hospital costs per case, respectively, versus nonhospitalist internists.[5,6]

As the field has matured, it may be that only the most talented and engaged internists have chosen to continue to provide care to their patients both in and out of the hospital. If so, then hospitalists may have replaced all but the best nonhospitalist general internists; this would make it more difficult for hospitalists to outperform their similarly trained general internist colleagues and explain the less vigorous hospitalist benefit in more recent studies. This would be acceptable if both parties were delivering the highest quality, safest, most cost-effective care. However, continued reports of unacceptable error rates and sky-rocketing healthcare costs offer continued opportunities for improvement; these opportunities will remain unmet without residency redesign toward more suitable hospitalist training.

CASE FOR SPECIALIZED TRAINING IN HOSPITAL MEDICINE: CLOSING THE TRAINING-PRACTICE GAP

The roles and responsibilities of a hospitalist have evolved over the past 15 years. Early hospitalists functioned essentially as internists in the hospital had for decades. The difference was that they focused their practice on the acute hospital setting, allowing their colleagues, often from the same group, to focus on the outpatient practice. However, they still provided the same type of care that they always did, that is, acute IM. Over time, hospitalists' roles have progressed beyond the reach of traditional IM training, and this has left newly minted hospitalists feeling unprepared to provide some services asked of them and consequently not well prepared to improve outcomes.[7]

Most notable is the advent of the comanagement model, which pairs a surgeon and a hospitalist in the care of a patient's surgical and medical issues, respectively. This places the hospitalist squarely in the middle of surgical care and outside the bounds of traditional IM training and practice. Although internists have always provided opinions regarding the medical care of surgical patients, this has traditionally been done in a consultative role, with internists offering advice in the form of recommendations and not managing the medical issue.[8] The comanagement model requires hospitalists to actively manage the patient alongside a surgical colleague, and this results in a level of responsibility sharing and communication foreign to the consultative model and rarely taught in training.[9] In many cases, hospitalists even serve as the attending of record for these patients; this arrangement frequently requires them to manage some or all of the surgical issues and necessitates skills that are not traditionally taught in IM training.

A recent article[10] examines a database of a national hospitalist corporation and offers a glimpse at how the hospitalist role has evolved away from traditional IM inpatient training. After reviewing the billing database for 141,354 patients seen by 436 community hospitalists in 133 hospitals in 11 cities, the authors concluded that nearly 30% of community hospitalist practice consists of underemphasized areas of traditional IM training, such as consultative medicine (6.4%), neurology (13.4%), orthopedics (6.4%), and general surgery (2.2%). Notably, only approximately 20% of the care provided to the orthopedic and general surgical patients was performed in the consultant role, highlighting the shift toward hospitalists acting as primary attendings on surgical patients. Furthermore, 46% of the patients were geriatric (>65 years old), with the largest subgroup of patients falling into the 75- to 84-year-old range. Overall, 5 of the top 15 diagnoses in this large community hospitalist practice involved neurologic or orthopedic disease states; this was a much different case mix than that seen in comparator studies of academic hospitalist groups, which likely reflect the types of patients that IM residents care for.

Plauth and colleagues[7] further evaluated the educational mismatches between IM training and hospital medicine practice. They surveyed 389 IM residency–trained hospitalists, a group that accounts for 86% of hospitalists,[11] regarding their level of preparedness to practice hospital medicine. They found significant educational deficiencies in many clinical areas deemed important to practice and underemphasized in training, such as geriatrics, neurology, palliative care, rehabilitation medicine, perioperative care, and consultative medicine. They also found important underemphasis in their understanding of healthcare economics as well as their level of preparation to carry out quality improvement and utilization review work and participate in multidisciplinary care teams. Further deficiencies were noted in the realms of ethics (eg, advanced directives, competency evaluation, and coordination of conservatorship), continuums of care (eg, coordination of care across settings, hospice care, functional status assessment, and prescription of appropriate levels of care), and communication skills (eg, communication with ancillary staff, referring physicians, and insurers). Finally, they found that IM training tended to overemphasize the procedural aspects of medicine, such as arterial line placement and pulmonary artery catheterization.

Additional survey data highlight the importance of the nonclinical aspects of the hospitalist job. In 2005, 59% of hospitalists reported working on utilization review projects, whereas 86% of groups partook in hospital quality improvement work, with 72% developing practice guidelines.[12] This represents a marked increase from 1999, when 53% and 46% of hospitalists reported involvement in quality improvement and practice guideline development, respectively, and further illustrates both the growing significance of this type of work to hospitalists and the shift away from traditional IM training and practice.[13] Furthermore, as frontline users of information systems, 54% of hospitalists are involved in electronic medical record and computerized provider order entry implementation.[12]

When taken together, these data help quantify the training-practice gap that exists for hospitalists trained through IM.[14] Furthermore, these findings help to at least partially explain why the field has not been able to fully achieve its promise. To realize this potential, hospitalists will need targeted training aimed at the core components that define the new specialty of hospital medicine. This can be done in a number of ways, but the most logical, efficient, and effective manner is IM residency redesign.

RESIDENCY REDESIGN: COMPONENTS OF HOSPITALIST-FOCUSED TRAINING

In recent years, there has been increasing pressure to redesign residency training to better prepare future internists for the rigors of practice. This emanates from a growing sense that the scope of IM is becoming too large, outstripping trainees' abilities to master the entire body of knowledge. In response, several influential IM professional groups have issued position papers calling for more vocationalized training that offers all trainees a broad core of IM complemented by training tailored to their future career path.[15–17] These more tailored offerings could include focused training in myriad areas, such as primary care medicine, women's health, human immunodeficiency virus care, and hospital medicine.

Training programs interested in adopting hospitalist-tailored training should commence the process by honestly assessing the current state of their program and examining the readiness of the local culture to accept this type of tailored approach.[18] This process will yield a sense of the current state of hospitalist training as well as the likelihood that changes in this direction will have broad institutional support. Once this programmatic assessment is complete, these findings should be evaluated against the Society of Hospital Medicine Core Competencies in Hospital Medicine as well as the areas of deficiency noted previously and summarized in Table 1.[19] These deficits should then be addressed through targeted educational initiatives.

Successful hospitalist-focused training should be built around the structure of the 6 core competencies of the Accreditation Council for Graduate Medical Education: patient care, medical knowledge, practice-based learning and improvement, interpersonal and communication skills, professionalism, and systems-based practice.[20] In many cases, this will mean strengthening and tailoring the curricular components that tie back to systems-based practice, communication skills, and practice-based learning and improvement. Additionally, this training should augment and not abandon or subjugate the important elements of subspecialty and primary care medicine. It is paramount that hospitalist trainees have a strong working knowledge of primary care medicine so that they remain well rounded, understand the capabilities of outpatient practice, and have a sound base should they later decide to practice in the outpatient environment.

Common areas of clinical deficiency include stroke care, palliative care, geriatric medicine, and consultative and perioperative medicine. Notably, perioperative medicine curricula should focus on the application of the Accreditation Council for Graduate Medical Education competencies of patient care, medical knowledge, and communication around the comanagement model of care, an area of significant undertraining and frequent clinical misadventure. Areas of nonclinical emphasis should include healthcare finance and the means by which it drives decision-making in hospital medicine, leadership skills, and continuums of care, especially effective transitions of care. Most importantly, future hospitalists should be trained in the tenets of hospital-based resource utilization and process and quality improvement. Optimally, this will result in a mentored project in which trainees develop, pilot, implement, and

Table 1. *Underemphasized Domains of Hospitalist Training**.

Non-clinical areas	Examples of specific training needs
Healthcare finance and economics	Recognizing the stakeholders who drive hospital decision making Understanding the role and effects of hospital utilization review Understanding the effect of hospital reimbursement on compensation
Patient safety and quality improvement	Common safety issues and analytical methods Role hospital systems play in medical errors, adverse events & patient safety Leading hospital-wide patient safety & QI initiatives Role in development, implementation and management of hospital information systems Role in practice guideline development & utilization
Continuums of care	Safeguarding transitions of care & patient hand-offs Appropriate use of the different types of postdischarge care
Practice management	Inpatient billing and coding Understanding terms and conditions of hospitalist employment contracts Appropriate medical error disclosure Understanding risk management Recognizing how to avoiding malpractice
Leadership	Means to effect institutional change Methods for negotiation & conflict resolution Use of personal influence to achieve goals Demonstrating the value of quality improvement and patient safety work
Communication	skills Appropriate and timely communication with providers and patients Organizing & facilitating patient family meetings Effectively delivering bad news Leadership & participation in multidisciplinary teams Medical error disclosure
Clinical	Examples of specific training needs
Orthopedics & trauma	Acute hip fracture care Management of joint arthroplasty Understanding acute rehabilitation medicine
Perioperative & consultative medicine	Preoperative risk assessment & patient optimization Perioperative venous thromboprophylaxis & diabetes management Postoperative pain management Assessment & management of the surgical abdomen Effective medical consultation & communication Differentiating the consultant & co-management role
Geriatric care	Appropriate medication use in the elderly Management of polypharmacy Limiting functional & nutritional decline in hospitalized patients Falls, delirium & decubitus ulcer prevention, recognition & treatment
Neurology	Management of blood pressure, diabetes and risk factor modification in acute stroke Indications and use of thrombolytics in acute stroke Risk assessment and secondary prevention of transient ischemic attack Acute seizure management Medical care of intracranial, subarachnoid & subdural hemorrhages
Palliative care & hospice	Acute pain & symptom management in the hospitalized patient Understanding the appropriate use and role of hospice benefits Appropriate feeding tube, nutrition and hydration use at the end of life

* Some of this underemphasized content might be appropriate for all residents

measure a hospital quality improvement or resource utilization project.

MECHANISMS TO DELIVER HOSPITALIST-FOCUSED TRAINING

Hospitalist-focused IM residency training can be delivered across a spectrum ranging from 1 general hospitalist elective to several specialized elective rotations to a comprehensive hospitalist track. The number and types of hospitalist offerings should be tailored to the individual training program on the basis of the current deficiencies and the level of institutional acceptance and commitment to vocational training. However, the known data suggest that US IM training programs have a significant chasm to cross in order to fill the hospitalist training-practice gap, and we recommend that programs consider a more comprehensive approach. This could include longitudinal offerings across all 3 years or more focused curricula in the second year and/or third year of training.

The decision to require this training of all residents versus only those interested in hospital-based careers, which could include more inpatient-focused subspecialties, should be driven by institutional interest in broad training versus tailored training. The benefits of offering 1 or more hospitalist electives is that they are easier to develop than a comprehensive track, allow residents to use their elective time to investigate areas of IM outside hospital medicine, and do not require residents to make an early residency decision about a hospitalist career. Furthermore, the elective approach allows other residents not interested in hospital medicine to sample some of the curricular components. The downside to this approach is that it is less likely to result in comprehensively trained hospitalists adroit at managing the myriad clinical and nonclinical aspects of a hospitalist career. On the other hand, a hospitalist track in which residents complete several hospitalist-focused "requirements" will more likely result in expertly trained hospitalists at the risk of undertraining in other areas of medicine. This latter risk is especially important in situations in which trainees opt for an alternate career.

Hospital medicine fellowships offer another postresidency option to address the training-practice gap. To date, the majority of hospitalist fellowships have focused on training hospitalist researchers. Still, several purely clinical fellowships exist for internal, family, and pediatric medicine residency graduates.[21] These options will likely grow in step with a greater awareness of the training deficits and a higher demand for focused training from hospitalist employers.[22]

SUMMARY

Hospital medicine offers an outstanding opportunity to improve the quality of the American healthcare system. However, the field has evolved rapidly and developed a clinical and nonclinical reach beyond the classic bounds of traditional IM practice, and the training offered in most traditional IM residency programs has not evolved to meet these needs. To fully achieve the needed improvements in resource utilization, patient safety, and quality improvement, the field of hospital medicine needs more focused training addressing the differences between the skills currently taught in IM training and the skills required for the practice of hospital medicine.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Wachter RM, Goldman L. The hospitalist movement 5 years later. *JAMA* 2002; 287: 487–494.
2. Meltzer D, Manning W, Morrison J, et al. Effects of physician experience on costs and outcomes on an academic general medicine service: results of a trial of hospitalists. *Ann Intern Med* 2002; 137: 866–874.
3. Auerbach AD, Wachter RM, Katz P, et al. Implementation of a voluntary hospitalist service at a community teaching hospital: improved clinical efficiency and patient outcomes. *Ann Intern Med* 2002; 137: 859–865.
4. Huddleston JM, Long KH, Naessens JM, et al. Medical and surgical comanagement after elective hip and knee arthroplasty. *Ann Intern Med* 2004; 141: 28–38.
5. Auerbach AD, Rasic MA, Sehgal N, et al. Opportunity missed: medical consultation, resource use, and quality of care of patients undergoing major surgery. *Arch Intern Med* 2007; 167: 2338–2344.
6. Lindenauer PK, Rothber MB, Pekow PS, et al. Outcomes of care by hospitalist, general internists, and family physicians. *N Engl J Med* 2007; 357: 2589–2600.
7. Plauth WH, Pantilat SZ, Wachter RM, et al. Hospitalists' perceptions of their residency training needs: results of a national survey. *Am J Med* 2001; 111: 247–254.
8. Goldman L, Lee T, Rudd P. Ten commandments for effective consultations. *Arch Intern Med* 1983; 143: 1753–1755.
9. Salerno SM, Hurst FP, Halvorson S, Mercado DL. Principles of effective consultation: an update for the 21st-century consultant. *Arch Intern Med* 2007; 167: 271–275.
10. Glasheen JJ, Epstein KR, Siegal E, et al. The spectrum of community-based hospitalist practice, a call to tailor internal medicine residency training. *Arch Intern Med* 2007; 167: 727–728.
11. Society of Hospital Medicine. 2007–2008 survey results. Available at: http://www.hospitalmedicine.org/am/template.cfm?section=home&template=/cm/contentdisplay.cfm&contentid=17184. Accessed April 2008.
12. Society of Hospital Medicine. 2005–2006 SHM survey: state of the hospital medicine movement. Available at: http://dev.hospitalmedicine.org/am/template.cfm?section=survey&template=/cm/contentdisplay.cfm&contentid=14352. Accessed April 2008.
13. Lindenauer PK, Pantilat SZ, Katz PP, Wachter RM. Hospitalists and the practice of inpatient medicine: results of a survey of the National Association of Inpatient Physicians. *Ann Intern Med* 1999; 130: 343–349.
14. Arora V, Guardiano S, Donaldson D, et al. Closing the gap between internal medicine training and practice: recommendations from recent graduates. *Am J Med* 2005; 118: 680–685.
15. Larson EB. Health care system chaos should spur innovation: summary of a report of the Society of General Internal Medicine Task Force on the Domain of General Internal Medicine. *Ann Intern Med* 2004; 140: 639–643.
16. Fitzgibbons JP, Bordley DR, Berkowitz L, et al. Redesigning residency training in internal medicine: a position paper from the Association of Program Directors in Internal Medicine. *Ann Intern Med* 2006; 144: 920–926.
17. Weinberger SE, Smith LG, Collier VU. Redesigning training for internal medicine. *Ann Intern Med* 2006; 144: 927–932.

18. Glasheen JJ, Siegal EM, Epstein K, et al. Fulfilling the promise of hospital medicine: tailoring internal medicine training to address hospitalists' needs. *J Gen Intern Med* 2008; 23: 1110–1115.
19. Pistoria MJ, Amin AN, Dressler DD, et al. The core competencies in hospital medicine. *J Hosp Med* 2006; 1(suppl 1): 1–95.
20. Accreditation Council for Graduate Medical Education. Core competencies. Available at: http://www.acgme.org/outcome/comp/compmin.asp. Accessed April 2008.
21. Ranji SR, Rosenman DJ, Amin AN, Kripalani, S. Hospital medicine fellowships: works in progress. *Am J Med* 2006; 119: 72.
22. Goodman PH, Januska A. Clinical hospital medicine fellowships: perspectives of employers, hospitalists, and medicine residents. *J Hosp Med* 2008; 3: 28–34.

Preoperative Evaluation of Hospitalized Patients

Stephen K. Sigworth, MD, MSHA

Mount Sinai School of Medicine, New York, NY

ABSTRACT

Carefully managing patients undergoing elective surgeries is difficult in the perioperative setting. However, this becomes increasingly complex in patients hospitalized for acute conditions that may or may not be related to the pending surgery. Not only must the consulting physician take into consideration any complications inherent to the surgical procedure, but must also consider all related comorbidities of the acute condition for which the patient was initially hospitalized plus any existing chronic conditions. A careful systematic approach should be undertaken in these circumstances, which consists of (1) perioperative risk stratification, (2) medical optimization, and (3) perioperative risk reduction. Risk stratification is determined by the patient's inherent perioperative cardiac risk factors, whereas medical optimization and risk reduction are actively determined during the hospital course. For perioperative risk stratification, the Revised Cardiac Risk Index is the simplest tool for accurately identifying those patients at increased perioperative risk for cardiac mortality and morbidity. Medical optimization involves performing any necessary preoperative testing that would help identify concurrent undiagnosed medical conditions that might require preoperative intervention or the initiation of certain medication regimens to optimize disease treatment. Lastly, perioperative risk reduction includes any modalities that would be started to decrease the risk of potential perioperative cardiac, pulmonary, or other surgery-related comorbidities. ***Mt Sinai J Med 75:442–448, 2008.*** © *2008 Mount Sinai School of Medicine*

Address Correspondence to:

Stephen K. Sigworth
Mount Sinai School of Medicine
New York, NY
Email: stephen.sigworth@
mountsinai.org

Key Words: perioperative medicine, cardiac evaluation, perioperative beta-blocker, surgery on hospitalized patient.

Preoperative evaluation is a challenging task for internists because of the myriad risks associated with the surgical period. Surgeons rely on the internist to conduct a thorough evaluation and optimize medical management, given the fact that the patient will soon be exposed to surgical risks superimposed upon their acute and chronic medical conditions.

The most important aspect of the preoperative evaluation is taking a comprehensive history that focuses on cardiopulmonary disease and a review of medications. If the patient has no cardiac history, assessing his functional capacity becomes paramount. Asking about activities such as exercise (walking, climbing stairs, and performing housework or activities of daily living) can reveal the extent of a patient's cardiac reserve. In addition, noting what limits exercise may lead to underlying, undiagnosed cardiac disease. Frequently, patients undergoing orthopedic or spinal surgeries have limited exercise because of various neurologic or musculoskeletal complaints. Inquiring whether exertion or other activities precede symptoms of chest pain or dyspnea may reveal early signs of coronary artery disease or congestive heart failure. These types of symptoms may require further diagnostic testing prior to any surgical procedures. Importantly, documenting a complete medication history, including over-the-counter medications and herbal medications, will highlight appropriate perioperative use or discontinuation of these particular medicines prior to the surgical procedure.

When a patient is being approached before noncardiac surgery, the following 3 aspects must be considered:

1. Perioperative risk stratification.
2. Medication optimization.
3. Perioperative risk reduction.

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20073

© 2008 Mount Sinai School of Medicine

It is the task of the internist to best approximate this risk, optimize medication management of all medical comorbidities, and offer a plan designed to reduce the risk of complications and prevent exacerbation of chronic disease. These 3 aspects need to be followed distinctly and separately to capture all of the necessary testing and evaluation prior to the surgical intervention to reduce the overall risk of complications during the surgical period.

RISK STRATIFICATION

Determining risk is the first step in the preoperative evaluation. A patient's risk is based on his medical history and cannot be changed. The patient can be medically optimized, but his risk remains the same, given his unique medical history.

Three cardiac risk indices have been widely disseminated. The first prospective study to determine variables predictive of cardiac complications was performed in 1977, and it was labeled the Cardiac Risk Index.[1] It identified 9 criteria that were associated with a major cardiac complication or death and assigned points based on these risks. The total number of points placed these patients in a certain risk class that was predictive of cardiac death or a major complication. In 1986, a second prospective study modified the Cardiac Risk Index to further delineate the severity of both coronary artery disease and congestive heart failure. The Modified Cardiac Risk Index[2] identified 13 criteria that were associated with a major cardiac complication or death and assigned points to each criterion based on the associated increase in risk. Patients were categorized into risk groups according to their number of points, which gave a likelihood ratio of developing a cardiac complication or death. It is important to note that prior risk indexes were not validated in other settings or populations yet became widely adopted.[3] In 1999, a third prospective study assessing cardiac risk, the Revised Cardiac Risk Index (RCRI),[4] identified only 6 criteria that were associated with cardiac complications or death (see Table 1). These include a history of coronary artery disease, congestive heart failure, cerebrovascular disease, chronic kidney disease, diabetes, and high-risk surgical procedures (intrathoracic, intraperitoneal, or major vascular surgery). This simpler index was not only derived but was also validated against a second cohort with an overall larger sample size. Interestingly, the smaller set of criteria in this index proved to have better diagnostic performance than the original Cardiac Risk Index, the Modified Cardiac Risk Index, and the American Society of Anesthesiology classification according to receiver operating characteristic analyses. As a result, the RCRI not only has become more widely used by practicing clinicians but also has been incorporated into subsequent studies assessing perioperative cardiac risk and usage of beta-blockers during the perioperative setting. The RCRI was also included in the algorithm of the 2007 American College of Cardiology (ACC)/American Heart Association (AHA) guideline on perioperative cardiovascular evaluation.

Table 1. *Revised Cardiac Risk Index.*

High-risk type of surgery
Ischemic heart disease
History of congestive heart failure
History of cerebrovascular disease
Insulin therapy for diabetes
Preoperative serum creatinine >2.0 mg/dL

Whichever risk index is used, one can determine a patient's perioperative cardiac morbidity and mortality. In choosing which risk index to use, most have used the RCRI because of its simplicity. In the hospitalized patient, by identifying any of the preexisting 5 medical conditions and having knowledge of the pending surgery, one can easily determine cardiac risk. It is important to remember that this is risk only and does not yet bring into consideration any management steps that can potentially reduce these risks.

CARDIAC EVALUATION

In September 2007, the ACC and AHA published their 2007 guidelines on perioperative cardiovascular evaluation and care for noncardiac surgery.[5,6] This marked a significant change from their prior guideline published in 2002. In the prior guideline, there was an algorithm[7] that helped direct a patient's need for preoperative cardiac testing prior to surgery. On the basis of a patient's clinical predictors and functional status, plus the general risk of the procedure, a noninvasive cardiac test was or was not recommended. If a noninvasive test was performed and had a negative result, the patient would be recommended to go to the operating room. A positive result from the noninvasive test would prompt further invasive cardiac testing. The updated ACC/AHA algorithm recommends only cardiac testing if it will change management. The new algorithm is much simpler and follows the strategy of assessing active cardiac conditions, functional capacity as derived by metabolic equivalents, the patient's clinical risk factors (as indicated in the RCRI), and the risk of surgery. It subsequently incorporates more intensive medical management to control the heart rate in the recommendations rather than proceeding to cardiac testing (which is reserved for patients with low functional capacity undergoing intermediate-risk surgery or vascular surgery). Although the algorithm should not rigidly dictate recommendations, it should be used as a guideline to help a physician determine whether further medical testing prior to surgery is required to better optimize patient outcomes.

With respect to invasive cardiac testing and revascularization prior to noncardiac surgery, a Veterans Affairs study in 2004 found that in patients with stable coronary disease, coronary revascularization prior to elective vascular surgery did not improve survival in comparison with medical management.[8] No significant differences with respect to negative postoperative outcomes occurred between the revascularized group and the nonrevascularized group, with the exception of a delay to surgery (54 versus 18 days). As a result, the practice of revascularization of diseased coronary arteries prior to surgery became less standard, with medical management of coronary disease during the surgical period a preferable option because of its less invasive nature.

In the hospitalized patient, there typically is a period of inactivity depending on how long the patient was sick prior to admission or how long the patient has been in the hospital. It is important to determine functionality prior to the patient becoming acutely ill, as cardiac disease is typically progressive, unless the patient was admitted with a cardiac cause.

When patients are hospitalized because of an acute cardiac condition, consultation from a cardiologist is often required to optimize patient management. For patients admitted with symptoms of congestive heart failure, a preoperative echocardiogram should be obtained. The results from this noninvasive test, while not adding iatrogenic risk, will inform the operative team of any cardiac contractility or valvular disorder and assist the anesthesiologist with proper perioperative fluid and blood pressure management. Importantly, any medication changes as a result of the findings from the echocardiogram can be initiated. If the patient has been admitted with symptoms of unstable coronary disease, a discussion with the consulting cardiologist should be considered to determine whether further intervention is warranted on the basis of the patient's particular symptoms and prior available test results.

PERIOPERATIVE BETA-BLOCKADE

Beta-blockers are antihypertensive medications that lower the blood pressure and heart rate. Through a reduction of the heart rate, a subsequent increase in diastolic perfusion occurs. The improved delivery of oxygen to myocytes lessens the overall potential for myocardial ischemia.[9] An added benefit is the antiarrhythmic activity (class II), which may decrease the risk of an unstable rhythm during surgery. Perioperative myocardial infarctions are often due to disruption of unstable coronary plaques and the resulting myocardial ischemia. Because beta-blockers decrease mechanical stress on these vessels through their reduction of the heart rate and contractility, ruptures due to mechanical causes are diminished.[10] Given the nature of the mechanism of action of these drugs, they can also lead to periods of hypotension and bradycardia, which are of concern while a patient is under general anesthesia. Both of these conditions can be medically managed, but poor outcomes may result from pronounced or prolonged episodes of these potential hazards.

Medical management of coronary disease has largely revolved around maintaining a low heart rate and thereby decreasing myocardial demand for oxygen. In a review of studies of perioperative beta blockade, patients receiving perioperative beta-blockers had significantly lower rates of myocardial ischemia, infarction, and cardiac deaths and were titrated to achieve target heart rates of less than 65.[11] The evidence that supported this was based on a number of studies that were small in size (<200 study subjects) and did not exhibit consistency with respect to the formulation of the beta-blocker, its dosage, or the timing of its use.[12–19] As a result of these studies, use of perioperative beta-blockers has been frequently prescribed for patients with multiple cardiac risk factors. The use of beta-blockers with respect to type, dosage, and duration became physician-directed as no clear guidelines, based on these small studies, were established.

Subsequent randomized controlled trials and meta-analyses have called into question the beneficial effect of beta-blockers during the surgical period.[20–23] In addition, a very large retrospective cohort study revealed that the use of beta-blockers in patients with 1 cardiac risk factor or fewer offers no benefit during the surgical period and may result in actual harm.[24] Therefore, the absolute use of beta-blockers during the perioperative period, regardless of cardiac risk, was called into question. However, the study did find that perioperative cardiac risk was reduced significantly for patients with 2 or more RCRI risk factors. As a result, beta-blocker use remains important for cardiac protection in these subgroups.[24] The ACC/AHA guidelines advocate perioperative beta-blockers for patients with coronary artery disease or high cardiac risk undergoing high-risk surgery or vascular surgery.[25] Currently, there are 2 large prospective, multicenter trials that should further delineate specific patient populations who will absolutely benefit from perioperative beta-blockade.

The heart of the acutely ill patient is typically under stress. Fever, pain, and other etiologies can increase the heart rate and subsequent myocardial demand for oxygen. If the patient has coronary disease or congestive heart failure, they are likely already on a beta-blocker, and titrating its dosage to achieve a more optimal heart rate and blood pressure should be done. However, if the patient is not already on a beta-blocker and has more than 1 cardiac risk factor as defined in the RCRI, the addition of a shorter-acting beta blocker can be initiated in the hospital setting as early as the risk factors are recognized. Finding the optimal dose prior to the patient going to the operating room should prevent unnecessary episodes of hypotension and bradycardia during the procedure.

PULMONARY EVALUATION

In 2006, the American College of Physicians published clinical guidelines for risk stratification for perioperative pulmonary complications.[26] Pulmonary complications were defined as atelectasis, pneumonia, respiratory failure, and exacerbation of underlying chronic lung disease. This comprehensive systematic review identified characteristics that predisposed the patient to potential pulmonary complications in the perioperative setting. These include age (>60; the risk increases with advancing age), chronic obstructive pulmonary disease, congestive heart failure, and functional dependence. Additionally, smoking history and delirium offer a moderate risk. Obstructive sleep apnea, although not an absolute risk factor for perioperative pulmonary complications, was noted to increase the risk of intensive care transfers and increased length of stay.[27] The type of surgery that the patient undergoes directly predisposes the pulmonary complication risk. Any surgeries that occur near the diaphragm, including abdominal, thoracic, and aortic surgeries, will increase the risk, as will neurosurgery, vascular surgery, and emergency surgery. Other procedural factors that contribute to pulmonary complications include surgery of a duration greater than 2.5 hours, likely because of prolonged periods of mechanical ventilation.[26]

Despite the recognition of predisposing factors that lead to pulmonary complications, it remains important to medically optimize patients with lung disease prior to the operating room. The pulmonary preoperative evaluation should always include a thorough review of the patient's pulmonary history and a physical examination of the lungs. Routine ordering of chest X-rays or pulmonary function tests was not found to significantly predict pulmonary complications.[26] Intervening in the postoperative period to encourage lung expansion is the mainstay for reducing the likelihood of atelectasis.[28] This can be achieved as follows: the hospitalized patient should sit upright instead of remaining in the supine position, and frequent incentive spirometry should be encouraged. In addition, adequate pain control will decrease splinting and allow the patient to have greater lung expansion. Effective use of beta-agonists and corticosteroids (inhaled, oral, or parenteral) in patients with mild bronchoconstriction remains the therapy of choice in hospitalized patients. For patients with severe bronchoconstriction, consultation with a pulmonologist will assist with management and ensure optimal medical therapies.

DIABETES MANAGEMENT

Achieving euglycemia in the acute illness phase, even in nondiabetic patients, is difficult because of ongoing infections, relative inactivity, surgery, and anesthesia.[29] Hyperglycemia can not only impair wound healing and decrease the body's ability to fight infection but also create an imbalance of pro-inflammatory and anti-inflammatory mediators that mal-affects the vascular endothelium and thrombosis.[29–31] It is well documented that strict glycemic control in the surgical intensive care setting decreases patient morbidity and mortality,[32–34] but this is less well defined in other settings. That said, maintaining relative glycemic control in the perioperative period should prove beneficial for the patient in preventing postsurgical complications.

The hospitalized setting for the patient undergoing surgery is a more controlled setting with respect to sugar management. The meals are defined, as are the delivery and timing of medication. Also, periodic evaluations of blood sugar are available to the consulting internist. Using long-acting or peakless insulin for delivery of a patient's basal insulin requirement, with the addition of short-acting or preprandial short-acting insulins, will improve glycemic control in the hospital. These insulins can be adjusted daily according to recorded blood sugars. If a patient is on oral antidiabetic agents with a history of good glycemic control, leaving him on his current medication should suffice if this is medically warranted and the patient maintains euglycemia. On the day of surgery, any preprandial or short-acting insulins and oral antidiabetic agents should be held as the patient is typically not allowed to eat. The long-acting or peakless insulins should be reduced as medically indicated to allow for decreased oral intake.

Maintaining a euglycemic state for the hospitalized patient during the perioperative setting is extremely important for optimizing patient outcomes. Patients should be discharged on their outpatient regimen upon discharge from the hospital if they are returned to their normal prehospitalized condition or on their new regimen if their condition and medications have been altered.

ADDRESSING ACUTE CONDITIONS

Surgery is best done electively with respect to the incidence of perioperative complications. For a patient who has been hospitalized, the goal of the consulting internist is to improve the urgent condition that necessitated hospitalization and to get the patient as close to his baseline as possible. The ideal situation would be to delay surgery until the acute condition is completely resolved. If the need for urgent or emergent surgical intervention outweighs continued medical management, then the patient will be at a higher risk for a perioperative complication because of the more emergent nature of the surgery and the patient's suboptimal baseline health status.

PREOPERATIVE TESTING

Typically, the hospitalized patient has multiple laboratory and radiology tests throughout his inpatient stay. As a result, there are likely few additional tests that will need to be ordered. Most institutions have specific testing requirements prior to surgery, and these can vary widely. They can include, but are not limited to, baseline electrolyte and kidney function tests, hemoglobin, pregnancy testing (population-specific), and electrocardiograms/chest X-rays (age-specific).

COMMUNICATION

This is largely the most important aspect with respect to preoperative evaluations. Clear communication with the surgeon and other consultants will more likely ensure that your recommendations will be followed, and this will result in improved outcomes for the patient. If possible, a discussion with the surgical team regarding not only the recommendations but also the timing of their initiation will ensure that the patient receives the necessary interventions in a timely fashion. Moreover, communication with the patient and the patient's family, with particular attention paid to comorbid conditions, can better prepare them for any perioperative complications should they arise. Most importantly, postoperative communication with the patient's primary care provider on discharge from the hospital, with any complications or changes in medication noted, will lessen the risk of transition errors and enable the patient to experience an uneventful recovery.

SUMMARY

Systematic preoperative evaluations can identify potential contributions to perioperative complications. The necessary testing is determined by comorbid conditions and current symptomatology. The adage that physicians "first, do no harm" is especially important in the preoperative setting. Any test that is ordered, especially any invasive test, should be ordered under the condition that the results of such a test will change the management of the patient prior to surgical intervention. There are certain tests that are used to create a baseline should complications arise (creatinine, hemoglobin, and electrocardiogram), but others should be performed only when symptoms warrant.

Optimization of medical management takes into consideration all medical conditions, medications (prescribed and purchased over the counter, including herbals), and symptoms. A careful assessment of the status of these conditions, plus optimization of medications to bring the patient's condition to the best possible state, will decrease the risk of perioperative complications. Appropriate use of beta-blockers for selected patients will reduce adverse cardiovascular events, and optimization of pulmonary function and glycemic control will improve recovery. Lastly, effective communication with all involved parties (surgeons, patients, families, and primary care providers) increases understanding of perioperative comorbidities and risks and ensures that all are prepared and informed should complications arise.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Goldman L, Caldera DL, Nussbaum SR, et al. Multifactorial index of cardiac risk in noncardiac surgical procedures. *N Engl J Med* 1977; 297: 845–850.
2. Detsky, AS, Abrams HB, McLaughlin , et al. Predicting cardiac complications in patients undergoing noncardiac surgery. *J Gen Intern Med* 1986; 1: 211–219.
3. American College of Physicians. Guidelines for assessing and managing the perioperative risk from coronary artery disease associated with major noncardiac surgery. *Ann Intern Med* 1997; 127: 309–312.
4. Lee TH, Marcantonio ER, Mangione CM, et al. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. *Circulation* 1999; 100: 1043–1049.
5. Fleisher LA, Beckman JA, Brown KA, et al. ACC/AHA 2007 guidelines on perioperative cardiovascular evaluation and care for noncardiac surgery: executive summary. *J Am Coll Cardiol* 2007; 50: 1707–1732.
6. Fleisher LA, Beckman JA, Brown KA, et al. ACC/AHA 2007 guidelines on perioperative cardiovascular evaluation and care for noncardiac surgery. *J Am Coll Cardiol* 2007; 50: e159–e241.
7. Eagle, KA, Berger PB, Calkins H, et al. ACC/AHA 2007 guidelines on perioperative cardiovascular evaluation and care for noncardiac surgery. *Circulation* 2002; 105: 1257–1267.
8. McFalls EO, Ward HB, Moritz TE, et al. Coronary-artery revascularization before elective major vascular surgery. *N Engl J Med* 2004; 351: 2795–2804.
9. London MJ, Zaugg M, Schaub MD, Spahn DR. Perioperative beta-adrenergic receptor blockade: physiologic foundations and clinical controversies. *Anesthesiology* 2004; 100: 170–175.
10. Maggio PM, Taheri PA. Perioperative issues: myocardial ischemia and protection–beta-blockade. *Surg Clin North Am* 2005; 85: 1091–1102.
11. Auerbach AD, Goldman L. Beta-blockers and reduction of cardiac events in noncardiac surgery: scientific review. *JAMA* 2002; 287: 1435–1444.
12. Urban MK, Markowitz SM, Gordon MA, et al. Postoperative prophylactic administration of beta-adrenergic blockers in patients at risk for myocardial ischemia. *Anesth Analg* 2000; 90: 1257–1261.
13. Zaugg M, Tagliente T, Lucchinetti E, et al. Beneficial effects from beta-adrenergic blockade in elderly patients undergoing noncardiac surgery. *Anesthesiology* 1999; 91: 1674–1686.
14. Poldermans D, Boersma E, Bax JJ, et al. The effect of bisoprolol on perioperative mortality and myocardial infarction in high-risk patients undergoing vascular surgery. Dutch echocardiographic cardiac risk evaluation applying stress echocardiography study group. *N Engl J Med* 1999; 341: 1789–1794.
15. Raby KE, Brull SJ, Timini F, et al. The effect of heart rate control on myocardial ischemia among high-risk patients after vascular surgery. *Anesth Analg* 1999; 88: 477–482.
16. Wallace A, Layug B, Tateo I, et al. Prophylactic atenolol reduces postoperative myocardial ischemia. McSPI Research Group. *Anesthesiology* 1998; 88: 7–17.
17. Mangano DT, Layug EL, Wallace A, et al. Effect of atenolol on mortality and cardiovascular morbidity after noncardiac surgery. Multicenter study of perioperative ischemia research group. *N Engl J Med* 1996; 335: 1713–1720.
18. Boersma E, Poldermans D, Bax JJ, et al. Predictors of cardiac events after major vascular surgery: role of clinical characteristics, dobutamine echocardiography, and beta-blocker therapy. *JAMA* 2001; 285: 1865–1873.
19. Stone JG, Foex P, Sear JW, et al. Risk of myocardial ischaemia during anaesthesia in treated and untreated hypertensive patients. *Br J Anaesth* 1988; 61: 675–679.
20. Juul AB, Wetterslev J, Gluud C, et al. Effect of perioperative beta blockade in patients with diabetes undergoing major non-cardiac surgery: randomised placebo controlled, blinded multicentre trial. *BMJ* 2006; 332: 1482.
21. Brady AR, Gibbs JS, Greenhalgh RM, et al. Perioperative beta-blockade (POBBLE) for patients undergoing infrarenal vascular surgery: results of a randomized double-blind controlled trial. *J Vasc Surg* 2005; 41: 602–609.
22. Yang H, Raymer K, Butler R, et al. The effects of perioperative beta-blockade: results of the Metoprolol after Vascular Surgery (MaVS) study, a randomized controlled trial. *Am Heart J* 2006; 152: 983–990.
23. Devereaux PJ, Beattie WS, Choi PT, et al. How strong is the evidence for the use of perioperative beta blockers in non-cardiac surgery? Systematic review and meta-analysis of randomised controlled trials. *BMJ* 2005; 331: 313–321.
24. Lindenauer PK, Pekow P, Wang K, et al. Perioperative beta-blocker therapy and mortality after major noncardiac surgery. *N Engl J Med* 2005; 353: 349–361.
25. Fleisher LA, Beckman JA, Brown KA, et al. ACC/AHA 2006 guideline update on perioperative cardiovascular evaluation for noncardiac surgery: focused update on perioperative beta-blocker therapy. *Circulation* 2006; 113: 2662–2674.
26. Smetana GW, Lawrence VA, Cornell JE. Preoperative pulmonary risk stratification for noncardiothoracic surgery: systematic review for the American College of Physicians. *Ann Intern Med* 2006; 144: 581–595.
27. Gupta RM, Parvizi J, Hanssen AD, Gay PC. Postoperative complications in patients with obstructive sleep apnea syndrome undergoing hip or knee replacement. *Mayo Clin Proc* 2001; 76: 897–905.
28. Bapoje SR, Whitaker JF, Schulz T, et al. Preoperative evaluation of the patient with pulmonary disease. *Chest* 2007; 13: 1637–1645.
29. Coursin DB, Connery LE, Ketzler JT. Perioperative diabetic and hyperglycemic management issues. *Crit Care Med* 2004; 32(suppl): s116–s125.
30. Marks JB. Perioperative management of diabetes. *Am Fam Phys* 2003; 67: 93–100.

31. Capes SE, Hunt D, Malmberg K, et al. Stress and hyperglycemia and prognosis of stroke in nondiabetic and diabetic patients. *Stroke* 2001; 32: 2426–2432.
32. Van den Berghe G, Wouters P, Weekers F, et al. Intensive insulin therapy in critically ill patients. *N Engl J Med* 2001; 345: 1359–1367.
33. Krinsley JS. Effect of an intensive glucose management protocol on the mortality of critically ill adult patients. *Mayo Clin Proc* 2004; 79: 992–1000.
34. Mesottten D, Van den Berghe G. Clinical potential of insulin therapy in critically ill patients. *Drugs* 2003; 63: 625–636.

DOI:10.1002/MSJ

Prevention of Venous Thromboembolism in Hospitalized Medical Patients

Joshua D. Lenchus, DO, RPh, and Amir K. Jaffer, MD

Division of Hospital Medicine, University of Miami, Miller School of Medicine, Miami, FL

ABSTRACT

The hospitalized medical patient in the US today has multiple risk factors for venous thromboembolism (VTE) and is therefore at a significant risk of developing this condition both during hospitalization and after discharge. The American College of Chest Physicians (ACCP) guidelines recommend pharmacologic prophylaxis unless there are contraindications. Low-molecular-weight heparins (LMWH) are the preferred class of drugs due to multiple advantages, including daily dosing and a decreased risk of heparin-induced thrombocytopenia, to name a few. Several emerging oral anticoagulants, including Factor Xa inhibitors and direct thrombin inhibitors, are currently being evaluated in Phase III clinical trials as prophylaxis against VTE in medically ill patients and may some day replace parenteral drugs. ***Mt Sinai J Med 75:449–453, 2008.*** *© 2008 Mount Sinai School of Medicine*

Key Words: Venous thromboembolism, low-molecular-weight heparin, prevention, hospitalized, medically ill.

In the United States, some 2 million to 5 million cases of deep vein thrombosis (DVT) occur annually. About 50% of these patients will experience a pulmonary embolus (PE) according to studies using ventilation-perfusion scanning; of these, about 250,000 will die. The spectrum of venous thromboembolism (VTE; DVT and PE) can vary from no symptoms to sudden death from PE.

Unrecognized PE is a leading cause of sudden death in hospitalized medical patients according to necropsy studies. Those that survive may suffer from sequelae, such as chronic pulmonary hypertension, which occurs in approximately 5% of PE cases, and post–thrombotic syndrome, which occurs in approximately 40% of patients with DVT at 2 years.[1]

More recent literature has spotlighted the outpatient arena; some studies have suggested that DVT occurs significantly more frequently (about 3-fold) in this setting as some 50% of patients with VTE were noted to have been hospitalized in the 3 previous months.[2] The explanation for this dramatic rise may be a single factor or some combination of factors. Hypotheses for inadequate in-hospital prevention include (1) absence of prophylaxis, (2) use of an unsuitable modality, (3) insufficient dose of the drug, and (4) ineffective duration of therapy. The DVT Free Registry underscored the issue of inadequate and omitted VTE prophylaxis for medically ill patients. This large, multicenter registry was conducted at 183 US hospitals, including more than 5000 people (inpatients and outpatients) with ultrasound-confirmed DVT. The percentage of medical inpatients who received prophylaxis in the 30 days prior to diagnosis was only 28%, which was lower than that of surgical patients (48%).[3] In a recent study, the International Medical Prevention Registry on Venous Thromboembolism (IMPROVE), only approximately 50% of hospitalized medical patients received prophylaxis.[4]

Multiple healthcare organizations such as the National Quality Forum, the Joint Commission on Accreditation of Healthcare Organizations, and the Agency for Healthcare Research and Quality have now identified VTE as a preventable condition in hospitalized patients; therefore, it is necessary that formal risk assessment be conducted as a first step, followed by the initiation of timely prophylaxis to improve patient safety.

Address Correspondence to:

Joshua D. Lenchus
University of Miami
Miller School of Medicine
Miami, FL
Email: jlenchus@med.miami.edu

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20074

© 2008 Mount Sinai School of Medicine

RISK ASSESSMENT

Virchow's triad encompasses the underlying etiologic factors for thrombosis: endothelial injury, stasis of blood flow, and hypercoagulability. Documented VTE risk factors reflect these underlying pathophysiologic processes. The more commonly recognized include (1) increased age, (2) prolonged immobility, (3) malignancy, (4) major surgery, (5) multiple traumas, (6) a history of VTE, and (7) chronic heart failure.[5] All risk factors are not given equal weight as the magnitude of risk conferred by these and others varies (Table 1). A cursory review of the aforementioned risk factors illustrates how imperative screening has become; a significant proportion of medically ill patients have 1 or more risk factors for VTE.

There are several tools used to assess and/or assign a patient's individual risk. However, the interaction of risk factors, as they relate to a specific patient's VTE risk, is currently unknown; there is evidence that one's risk increases proportionally to the number of predisposing factors present.

EVIDENCE AND OPTIONS FOR PROPHYLAXIS

Mechanical Prophylaxis

Mechanical forms of prophylaxis include graduated compression stockings, intermittent pneumatic compression devices, and venous foot pumps. Graduated compression stockings have been evaluated in patients with stroke and myocardial infarction. Intermittent pneumatic compression devices and venous foot pumps have not been studied in randomized controlled trials in general medical patients. Although there are data supporting the efficacy of these devices in surgical patients, the American College of Chest Physicians guidelines recommend against the use of mechanical forms of prophylaxis in medical patients unless there is a contraindication to pharmacological prophylaxis.[6]

Unfractionated Heparin

Unfractionated heparin (UFH) is a heterogeneous mixture of repeating polysaccharide chains of various sizes with an average size of approximately 15000 Da. It binds anti-thrombin III and facilitates anti-thrombin III–mediated inactivation of factors IIa, IXa, Xa, and XIIa; of these, IIa and Xa are most responsive to inhibition. Because of its large size, UFH is only partially absorbed from subcutaneous tissue, and it has a variable anticoagulant response because of interactions with plasma proteins, macrophages, and endothelial cells.[7] However, in prophylactic subcutaneous doses (5000 units 2 or 3 times daily), monitoring of the activated partial thromboplastin time is not required; however, in some cases, (eg, frail or elderly patients), prophylactic subcutaneous doses may slightly prolong the activated partial thromboplastin time. UFH also binds to platelets and platelet factor 4 and may precipitate heparin-induced thrombocytopenia (HIT). At least 3 clinical trials[8–10] have compared the efficacy of subcutaneous UFH with placebo and found that prophylactic doses of UFH decrease the relative risk (RR) of DVT, as detected by fibrinogen uptake testing, by about 70% without increasing the risk of bleeding.[6]

Low-Molecular-Weight Heparins

LMWHs are derived from UFH through a chemical depolymerization, or fractionation, process. About one-third the size of UFH, they have a molecular weight of approximately 5000 Da. These smaller molecules have superior bioavailability, being readily absorbed from the subcutaneous tissue, and thus elicit a more predictable anticoagulant response than UFH. In contrast to UFH, LMWHs have only minimal nonspecific binding to plasma proteins, endothelial cells, and monocytes,[7] which results in a predictable dose response; this obviates the need for laboratory monitoring, even when used in full, therapeutic, weight-based dosing. Balanced against UFH, the LMWH agents (1) have a longer plasma half-life, which allows them to be dosed subcutaneously once or twice daily, (2) do not bind to platelets as readily as UFH, and (3) carry a lower risk of HIT. Because of their smaller size, LMWHs tend to preferentially inhibit factor Xa, whereas UFH tends to inhibit both factors Xa and IIa equally.[7]

LMWHs have been evaluated in 2 large, placebo-controlled clinical trials for the prevention of VTE in hospitalized medical patients. In the first, prophylaxis in Medical Patients with Enoxaparin (MEDENOX), almost 50% of the patients were older than 75 years (mean age ~ 73 years). Inclusion criteria were New York Heart Association class 3 or 4 heart failure, acute respiratory failure without mechanical ventilation, acute infection without septic shock, acute rheumatic disease, and inflammatory bowel disease. The primary outcome was assessed in 866 patients. Enoxaparin (40 mg subcutaneously once daily) decreased the rate of VTE from 15% to 5% ($P = 0.0002$) without any concomitant increase in the rate of bleeding or thrombocytopenia in comparison with placebo.[11] Enoxaparin is Food

DOI:10.1002/MSJ

Table 1. *Risk Factors for VTE in Medically Ill Patients.*

Immobility	Inflammatory bowel disease
Malignancy	Nephrotic syndrome
Cancer therapy	Myeloproliferative disorders
Radiotherapy, chemotherapy, or hormonal therapy	Estrogen-containing oral contraception, HRT, or SERMs
Previous VTE	Obesity
Increased age	Smoking
Pregnancy and postpartum period	Varicose veins
Decompensated heart failure	Respiratory failure
Infections (eg, UTI and pneumonia)	Central venous catheterization
Sepsis	Inherited or acquired thrombophilia

NOTE: This table was modified from Geerts *et al.*[6] **Abbreviations:** HRT, hormone replacement therapy; SERMS, selective estrogen-receptor modulator; UTI, urinary tract infection; VTE, venous thromboembolism.

Table 2. *Various Pharmacologic Agents for Medical Prophylaxis.*

	UFH	Enoxaparin	Dalteparin	Fondaparinux
Class	UFH	LMWH	LMWH	Synthetic pentasaccharide, factor Xa inhibitor
FDA-approved for medical prophylaxis	Yes	Yes	Yes	No
Approved dose	5000 units SC every 8 hours (3 times daily)	40 mg SC once daily	5000 IU SC once daily	2.5 mg SC once daily*
Dosing in renal insufficiency (CrCl: 10–30 mL/minute)	Yes: consider 5000 units SC twice daily	Yes: adjust to 30 mg SC once daily	Not indicated	Contraindicated
Reversal agent?	Protamine sulfate (complete reversal)	Protamine sulfate (60% – 75% reversal)	Protamine sulfate (60% – 75% reversal)	Recombinant factor VIIa?

Abbreviations: CrCl, creatinine clearance; FDA, Food and Drug Administration; LMWH, low-molecular-weight heparin; SC, subcutaneously; UFH, unfractionated heparin.
* Fondaparinux is not FDA-approved for medical prophylaxis

and Drug Administration (FDA)–approved for VTE prophylaxis in medically ill patients.

The second, Prevention of Recurrent Venous Thromboembolism (PREVENT), was an international, multicenter, randomized, double-blind, placebo-controlled trial evaluating dalteparin's efficacy and safety. The inclusion criteria were similar to those of the MEDENOX study as they encompassed patient with acute congestive heart failure, non–ventilator-requiring acute respiratory failure, infection without septic shock, acute rheumatologic disorders, and inflammatory bowel disease. The primary outcome was incidence of VTE and sudden death at day 21; this was assessed in 2991 patients. Dalteparin (5000 IU) decreased the rate of VTE, as detected by compression ultrasound, from 5% in the placebo group to 2.8%; this was an RR reduction of 45% ($P = 0.0015$). The authors concluded that the use of dalteparin reduced the incidence of VTE without increased risk of bleeding.[12] Dalteparin is FDA-approved for VTE prophylaxis in medically ill patients.

At least 2 meta-analyses have evaluated LMWHs versus UFH. In the first, which included approximately 4600 patients enrolled in 9 trials, LMWH was compared to UFH. No significant effect was observed on DVT, clinical PE, or mortality; however LMWH reduced the risk of major hemorrhage by 52% ($P = 0.049$).[13] In a more recent meta-analysis, 36 studies were included, comparing placebo to UFH and LMWH. Both agents were associated with a reduced risk of VTE. A UFH dosage of 5000 units 3 times daily was more effective in preventing DVT than a dosage of 5000 units twice daily in comparison with the control [RR, 0.27, and 95% confidence interval (CI), 0.20–0.36, versus RR, 0.52, and 95% CI, 0.28–0.96]. Neither UFH nor LMWH reduced mortality. When directly compared with UFH, LMWH was associated with a lower risk of DVT (RR, 0.68, and 95% CI, 0.52–0.88) and injection-site hematoma (RR, 0.47, and 95% CI, 0.36–0.62), but no difference was seen between the 2 agents with respect to risk of bleeding or thrombocytopenia. This latter analysis clearly illustrates that either agent reduces VTE risk in hospitalized medical patients, but neither alters mortality. In a direct comparison, LMWH is more effective in preventing DVT.[14]

Synthetic Pentasaccharide (Fondaparinux)

Fondaparinux is a synthetic analogue of the pentasaccharide sequence that mediates the interaction of heparin with antithrombin. It inhibits both free and platelet-bound factor Xa. It binds antithrombin with high affinity, has close to 100% bioavailability, and has a plasma half-life of 17 hours, which permits once daily administration. The drug is excreted unchanged in the urine and, therefore, is contraindicated in patients with severe renal impairment (eg, creatinine clearance < 30 mL/minute). It does not bind platelet factor 4 in vitro and thus should not cause HIT. Fondaparinux has been evaluated in hospitalized medical patients in a randomized clinical trial [Arixtra for Thromboembolism Prevention in a Medical Indications Study (ARTEMIS)]. Fondaparinux (2.5 mg subcutaneously once daily) decreased the rate of venographically confirmed DVT from 10.5% to 5.6% ($P = 0.029$); there was also a decrease in fatal PE from 1.7% to 0.7% ($P = 0.029$). A reduction in overall mortality from 6% to 3.3% ($P =$ not significant) was observed. Major bleeding occurred in 0.2% of patients in both groups.[15] Fondaparinux is not FDA-approved for VTE prophylaxis in medically ill patients.

Oral Anticoagulants

Oral anticoagulants [eg, vitamin K antagonists (VKAs)] have not been adequately studied in hospitalized medical patients. As VKAs take several days to achieve therapeutic anticoagulation, we do not recommend using them de novo as VTE prophylaxis. However, patients taking an oral VKA in the outpatient setting, who have a therapeutic international normalized ratio during hospitalization, are probably adequately protected from VTE and do not need additional pharmacological prophylaxis.

Emerging Anticoagulants

Some of the newer anticoagulants are still under phase III testing for prevention of VTE in medically ill patients; these include oral direct thrombin inhibitors and anti-Xa inhibitors.

DURATION OF THERAPY

Most clinical trials have used prophylaxis from 7 to 14 days. The Extended Clinical Prophylaxis in Acutely Ill Medical Patients (EXCLAIM) trial evaluated duration of therapy. The results were presented recently at the 21st Congress of the International Society on Thrombosis and Haemostasis: 5105 patients met inclusion criteria, 5049 received open-label prophylaxis with enoxaparin (10±4 days), 2013 patients subsequently received extended-duration enoxaparin, and 2027 received placebo (each for 28 ± 4 days). Subjects were at least 40 years of age, had been recently immobilized (for at least 3 days), and had a predefined acute medical illness. Mobility was defined as level 1 (total bed rest or sedentary) or level 2 (bathroom privileges). After a planned blinded interim analysis, the trial's data safety monitoring board recommended an amendment to changing the criteria of level 2 mobility to include age>75 years, prior VTE, diagnosed cancer, or a combination of these.

When compared with placebo, extended-duration enoxaparin following the open-label, standard-duration enoxaparin reduced the RR of VTE by 44% (2.8% versus 4.9%, $P = 0.0011$). There was major bleeding in 12 subjects who received extended-duration enoxaparin and in 3 subjects receiving placebo (0.6% versus 0.1% $P = 0.0192$). There was no difference in all-cause mortality between the extended-duration enoxaparin and placebo groups at 6 months (10.1% versus 8.9%, $P = 0.179$). The authors concluded that 38 days, or approximately 5 weeks, of enoxaparin (40 mg subcutaneously once daily) significantly reduced the overall incidence of VTE versus a 10-day regimen in acutely ill medical patients with reduced mobility. This reduction in overall risk for VTE was consistent with a reduction in risk for asymptomatic proximal DVT and symptomatic VTE. On the basis of this trial, we expect that 5 weeks of prophylaxis for those over 75 years of age, with a prior history of VTE, or with active cancer will become the standard of care. In the absence of these criteria, we support the notion of up to 14 days of prophylactic therapy.

CONCLUSIONS

All hospitalized medical patients should undergo an initial assessment of VTE risk factors and should be prescribed pharmacologic prophylaxis with LMWH for 7 to 14 days in the absence of any contraindications to anticoagulants. If no pharmacological prophylaxis is warranted due to the absence of risk factors, we recommend reassessing the patient daily, or when the clinical condition changes, and instituting such preventive therapy at that time. If contraindications to such prophylaxis exist, then mechanical methods should be employed.

DISCLOSURES

Potential conflict of interest: Amirk Jaffer is a Consultant for Sanofi-Aventis and Astra Zeneca. Amir Jaffer is in the Speaker's Bureau for Sanofi-Aventis. Amir Jaffer is also a Board Member of the AF-Forum and SPA QI.

REFERENCES

1. Hirsh J, Hoak J. Management of deep vein thrombosis and pulmonary embolism. A statement for healthcare professionals. Council on Thrombosis (in consultation with the Council on Cardiovascular Radiology), American Heart Association. *Circulation* 1996; 93: 2212–2245.
2. Spencer FA, Lessard D, Emery C, et al. Venous thromboembolism in the outpatient setting. *Arch Intern Med* 2007; 167: 1471–1475.
3. Goldhaber SZ, Tapson VF. A prospective registry of 5,451 patients with ultrasound-confirmed deep vein thrombosis. *Am J Cardiol* 2004; 93: 259–262.
4. Tapson VF, Decousus H, Pini M, et al. Venous thromboembolism prophylaxis in acutely ill hospitalized medical patients: findings from the International Medical Prevention Registry on Venous Thromboembolism. *Chest* 2007; 132: 936–945.
5. Anderson FA Jr, Spencer FA. Risk factors for venous thromboembolism. *Circulation* 2003; 107(suppl 1): I9–I16.
6. Geerts WH, Bergquist D, Pineo GF, et al. Prevention of venous thromboembolism: American College of Chest Physicians evidence-based clinical practive guidelines (8th Edition). *Chest* 2008; 133: 381–453.
7. Hirsh J, Raschke R. Heparin and low-molecular-weight heparin: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest* 2004; 126(():suppl): 188S–203S.
8. Cade JF, Andrews JT, Stubbs AE. Comparison of sodium and calcium heparin in prevention of venous thromboembolism. *Aust N Z J Med* 1982; 12: 501–504.
9. Gallus AS, Hirsh J, Tutle RJ, et al. Small subcutaneous doses of heparin in prevention of venous thrombosis. *N Engl J Med* 1973; 288: 545–551.
10. Belch JJ, Lowe GD, Ward AG, et al. Prevention of deep vein thrombosis in medical patients by low-dose heparin. *Scott Med J* 1981; 26: 115–117.
11. Samama MM, Cohen AT, Darmon JY, et al. A comparison of enoxaparin with placebo for the prevention of venous thromboembolism in acutely ill medical patients. Prophylaxis in Medical Patients with Enoxaparin Study Group. *N Engl J Med* 1999; 341: 793–800.
12. Leizorovicz A, Cohen AT, Turpie AG, et al. Randomized, placebo-controlled trial of dalteparin for the prevention of venous thromboembolism in acutely ill medical patients. *Circulation* 2004; 110: 874–879.
13. Mismetti P, Laporte-Simitsidis S, Tardy B, et al. Prevention of venous thromboembolism in internal medicine with unfractionated or low-molecular-weight heparins: a meta-analysis of randomised clinical trials. *Thromb Haemost* 2000; 83: 14–19.
14. Wein L, Wein S, Haas SJ, et al. Pharmacological venous thromboembolism prophylaxis in hospitalized medical patients: a meta-analysis of randomized controlled trials. *Arch Intern Med* 2007; 167: 1476–1486.
15. Cohen AT, Davidson BL, Gallus AS, et al. efficacy and safety of fondaparinux for the prevention of venous thromboembolism in older acute medical patients: randomised placebo controlled trial. *BMJ* 2006; 332: 325–329.

DOI:10.1002/MSJ

Impact of Obesity on Hospitalized Patients

Brian Markoff, MD, and Alison Amsterdam, MD, FACP

Division of General Internal Medicine, Department of Medicine, Mount Sinai Medical Center, New York, NY

ABSTRACT

Obesity is increasing at an alarming rate worldwide and is a risk factor for cardiac disease, diabetes, and stroke. This provides a challenge for the physicians caring for this patient population in the hospital. Skin integrity, medication dosing, testing, and nutrition are all altered by obesity. We summarize some of the current data on caring for the obese inpatient. Unfortunately, few data on this unique inpatient population exist. ***Mt Sinai J Med 75:454–459, 2008.*** © *2008 Mount Sinai School of Medicine*

Key Words: hospitalized patient, obesity, metabolism, mobility, nutrition.

In 2005, the World Health Organization estimated that over 400 million adults (approximately 7%) suffer from obesity worldwide, which is defined as a body mass index of ≥30 kg/m^2 (Table 1). Current projections forecast that this number will almost double by the year 2015.[1] Because obesity is a risk factor for multiple diseases including cardiac disease, stroke, diabetes, and certain malignancies,[2] there is a potential for increased numbers of hospitalizations and healthcare costs.[3,4] In addition, obese patients are at higher risk for inpatient complications unrelated to their presenting illness due to difficulties in mobility, ventilation, tissue perfusion, nutrition, and pharmacokinetics.[5–10] However, few data exist on how to care for the unique needs of these patients while they are hospitalized. Thus, our aim is to review current challenges and recommendations for inpatient management of this patient population.

Address Correspondence to:

Brian Markoff
Mount Sinai Medical Center
New York, NY
Email: brian.markoff@
mountsinai.org

SKIN CARE AND MOBILITY

Decreased mobility is known to contribute to ulcerations in areas that are exposed to increased and constant pressure. Obese patients can have ulcers in atypical areas, such as skin folds and around catheters, through the same mechanisms.[11] The Braden Scale for Predicting Pressure Sore Risk is an objective tool that can identify patients at risk upon admission.[12] This allows staff to enact prophylactic measures to prevent skin breakdown such as frequent repositioning, limb elevation, limited bathing of dry skin, frequent application of lotions and (when appropriate) antifungals, and use of wrinkle-free sheets to prevent lymphedema and/or skin necrosis.[13–15] Trapezes and Hoyer lifts can assist staff in achieving these goals.

Early identification of existing wounds is also important to prevent progression. Frequently changed absorbent dressings, adequate nutrition, avoidance of unnecessary adhesives and pressure dressings, and the use of skin ointments and sealants can promote wound healing.[16] Addressing barriers that may lead to incontinence can reduce moisture in pressure-prone areas such as the sacrum and buttocks.[17] Providing oversized walkers and bedside commodes promotes mobility and even increases self-esteem.[18] Stocking properly fitting diapers made with breathable cloth outer layers and having suppliers who can provide low-air-loss mattresses are options for patients who are severely incontinent and/or are unable to ambulate.[19] Finally, having variously sized ostomy supplies can help decrease the risk of stool leakage.[20,21]

PHARMACOKINETICS

People with obesity have different proportions of lean mass to fat than their peers. This may lead to alterations in cardiac performance, blood flow, drug binding, cytochrome P450 metabolism, and renal excretion of drugs. These factors can lead to variations in drug dosing, as outlined next.

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20072

© 2008 Mount Sinai School of Medicine

Table 1. *World Health Organization Classification of Body Mass Index and Obesity.*

Classification	Body Mass Index (kg/m^2)	Grade
Underweight	<18.5	n/a
Healthy	18.5–24.5	0
Overweight	25.0–30.0	1
Obese	30.0–40.0	2
Very obese	>40.0	3

Abbreviation: n/a, not applicable.

Antihypertensives

Lifestyle modification is the primary intervention for the treatment of hypertension in patients without diabetes or risk factors for coronary disease.[22] If conservative measures are unsuccessful, diuretics are the first line of agents suggested for management. In the past, these agents were avoided in obese patients despite their ability to decrease intravascular volume because they were thought to increase the incidence of insulin resistance. More recently, large-scale follow-up studies have shown no effect on the risk of developing type 2 diabetes,[23] and this has made them a more popular choice for this population.

For patients with an increased risk of coronary disease, beta-blockers have been shown to reduce morbidity and mortality in patients with primary hypertension.[22] Obese patients have an increased risk of left ventricular hypertrophy, sudden death, and congestive heart failure, and there is a suggestion of increased sympathetic tone.[24,25] Thus, beta-blockers may be an even more attractive choice for blood pressure control in this population. However, because older agents in this class have been shown to increase weight and alter carbohydrate metabolism,[26] close follow-up to watch for these complications should occur even after patients are discharged from the hospital. Using newer agents, such as carvedilol, may decrease these risks, but more studies are needed to confirm this finding. If patients experience weight gain, an alternative is the non-dihydropyridine calcium channel blocker verapamil. By inhibiting calcium influx into cardiac smooth muscle cells, verapamil decreases peripheral vascular resistance and can decrease heart rate. However, unlike beta-blockers, verapamil has not been found to have undesirable metabolic consequences.[27] There are currently minimal or no data regarding weight gain and efficacy for other calcium channel blockers in the treatment of hypertension in this patient population.

Because obese patients are at risk for developing diabetes, microalbuminuria, and left ventricular hypertrophy, angiotensin converting enzyme inhibitors and angiotensin receptor blockers should also be considered for obesity-related hypertension. Both classes of agents have been shown to decrease the risk of developing insulin-dependent diabetes, improve poor renal function, and promote cardiac remodeling.[28–30] In addition, it has been shown that adipose tissue plays a role in the renin-angiotensin axis in the regulation of body weight, so these agents are potentially beneficial in obese patients by promoting weight loss.[31,32]

Antimicrobials

Traditionally, antimicrobial dosing has been based on safety, efficacy, and minimum inhibitory concentrations.[33] However, these calculations have often not taken into account the differences in drug distribution that are seen in obese patients because of differing tissue to body fluid ratios. [34] Certain considerations should be made when several commonly used antimicrobials are being dosed.[35] For example, when vancomycin and daptomycin are being dosed, the total body weight should be used. For aminoglycosides and antifungals, adjusted and ideal body weights, respectively, are preferable. Beta-lactams, fluoroquinolones, and several carbapenems may require higher than recommended doses to achieve minimum inhibitory concentrations in obese patients. Finally, the dosing frequency of vancomycin, aminoglycosides, and daptomycin should be based on renal function and drug levels.[36] It is noteworthy that the majority of the data supporting these alterations are from small, case-based studies.

Anticoagulants

Decreased mobility and obesity are risk factors for venous thromboembolism.[37,38] However, there are minimal data regarding anticoagulants in the obese inpatient population. What is available is the result of bariatric surgical studies with small sample sizes. Low-molecular-weight heparins have different dosage curves and times to maximum anti-Xa levels for obese patients.[9,39] Therefore, dosing low-molecular-weight heparins by total or adjusted body weight rather than lean body mass has been recommended because absorption is affected by the quantity of adipose tissue.[10] Using this adjustment, low-molecular-weight heparins have been shown to be effective in reducing venous thromboembolism in obese patients in surgical studies.[40] Similar recommendations have been made regarding the dosing of subcutaneous unfractionated heparin because of differences in the quantity of adipose tissue.[41,42] Dosing protocols using anti–factor Xa levels are available, but again they are based on small studies.[43]

Oral Hypoglycemics and Insulin

Because most patients with non–insulin-dependent diabetes have at one time been treated with an oral hypoglycemic and up to 70% of diabetics are obese, pharmacokinetic differences in this patient population are imperative to the successful management of the disease.[44] Unfortunately, conflicting results have been demonstrated within classes of medications, likely because of the small sample sizes used. For example, a study of 12 weeks of glyburide in 20 patients with adult-onset diabetes showed that obese patients may be more sensitive to the agent.[45] Conversely, separate studies of glipizide in 12 diabetics[46] and glimeperide in 14 matched diabetics demonstrated no difference in the metabolism of either agent.[47] Data for short-acting insulins show similar inconsistencies and study design limitations.[48] Unfortunately, the majority of agents within both previously mentioned classes of drugs can lead to weight gain, perpetuating diabetes and the need for these medications. The exception is metformin. Multiple studies have shown that this agent, which acts by inhibiting gluconeogenesis, can lead to weight loss in diabetic patients, and this makes it an attractive medication choice for many type 2 diabetics. For the hospitalized patient, caution should be exercised if metformin is continued during the inpatient period. This agent has been shown to contribute to acute kidney failure and lactic acidosis if it is not discontinued prior to radiographic procedures involving intravenous contrast and other interventions that have the potential to decrease renal function. Thus, it is strongly recommended that the agent be stopped at the time of the procedure and reinstated 48 hours after its completion if renal function is within normal limits.

TESTS AND PROCEDURES

The obese patient presents multiple challenges from a diagnostic perspective. Not only does the patient's body habitus inhibit an accurate physical examination,[49] but it also may limit the types and quality of radiologic examinations and other procedures. Recent articles have documented the increase in radiology reports that include body habitus as a limiting factor in the quality of the examination.[50] The quality of plain films and ultrasounds is inferior with obese patients.[51] Multiple factors such as the weight limit of the tables and the aperture diameter of the apparatus being used make certain procedures impossible, including fluoroscopy, computerized tomography, magnetic resonance imaging, and nuclear scans.[52] Even transportation to and from the examination suites can be limited by the availability of equipment and personnel suitable for the obese patient. Obese patients may also need increased amounts of intravenous contrast to obtain appropriate images, and intravenous access can be difficult in some patients. Finally, obese patients present several challenges for interventional radiologists, including sedation issues, positioning issues, adequate instrument length, and lesion targeting issues.

Despite all this, some studies have shown that although images may be inferior, the diagnostic yield of certain tests is equivalent between normal-weight and obese patients.[53] Other diagnostic testing is also altered by body habitus. For example, cardiac catheterization is more difficult in the severely obese for many of the reasons mentioned previously (fluoroscopy, table weight, and positioning issues) and because of problems accessing the femoral vessels and hemostasis issues post-procedure. Alternative approaches, such as the use of upper extremity vessels for access, have been used in such cases.[54,55] Pulmonary function tests are also altered by obesity because of a decrease in chest compliance and an increase in airway resistance. This can lead to decreased forced vital capacity, forced expiratory volume in 1 second, total lung capacity, and functional residual volume and increased residual volume. Diffusing capacity should remain normal.[56] Despite these findings, 1 small study reported no correlation between abnormal forced vital capacity, forced expiratory volume in 1 second, and mid-flow rates and pulmonary complications after gastric bypass surgery.[57] Hemodialysis access can also be adversely affected by obesity. Compression of the venous outflow of the upper extremity due to increased body fat has been shown to contribute to premature hemodialysis vascular access failures.[58] Unfortunately, this cannot be predicted by preoperative vein mapping because obese and nonobese patients have similar rates of arteriovenous fistula functioning.[59]

NUTRITION

Nutritional status in obese inpatients can be challenging to assess because of the difficulty in obtaining appropriate equipment for anthropometric assessment, inaccurate calorie counts, and differences in provider assessments. The patient's weight history, nutritional status before hospitalization, and diet history are all important factors in determining the patient's correct nutritional needs during the inpatient stay.[60] Patients with moderate-to-severe obesity may benefit from hypocaloric feeding, which is not recommended for malnourished patients. This and other feeding strategies can maintain nitrogen balance even in critically ill patients while maintaining glycemic control and preventing overfeeding.[61] The caloric deficit of a moderately to severely obese patient should not be above 500 to 1000 kcal/day.[62] Recommendations for mildly obese patients suggest weight-maintenance feedings until the acute illness has resolved.[62] All obese patients should be given diets appropriate to the needs of their comorbid medical conditions, and this is where a multidisciplinary approach including a registered dietitian may prove beneficial. As previously mentioned, enteral and parenteral access for feeds can be challenging in obese inpatients secondary to imaging issues, instrumentation issues, and venous access issues.[63]

DISCHARGE

The obese patient may pose continued challenges at the time of discharge. Because of mobility issues, obese patients are more likely to have longer lengths of stay,[64] higher costs of hospitalization, and increased need for professional assistance and often require discharge to an inpatient rehabilitation center after orthopedic procedures.[65–67] Similar results have been noted in the evaluation of obese patients who required critical care stays or those who were diagnosed with obesity-associated hypoventilation syndromes.[6,68] Hospitals, transport companies, and families must also have the staff and equipment to safely transport this population from the hospital to the discharge destinations. Home care must be carefully coordinated prior to discharge so that all necessary equipment and plans of care are in place to minimize immobility and improve quality of care after hospitalization.[19,69]

CONCLUSIONS

The obesity epidemic is responsible for more obese patients needing hospitalization then ever before. This population presents many challenges for the treating providers and institutions. Despite the great numbers of such patients, there is a paucity of quality literature on how to effectively care for this important inpatient population. Future areas of research include trials to address the question of which agents are best for treating cardiac and metabolic disturbances in this patient population as well as quality control issues related to inpatient care and discharge.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Obesity and overweight. Available at: http://www.who.int/mediacentre/factsheets/fs311/en. Accessed January 2008.
2. Wannamethee G, Shaper AG, Walker M. Overweight and obesity and weight change in middle aged men: impact on cardiovascular disease and diabetes. *J Epidemiol Community Health* 2005; 59: 134–139.
3. Thompason D, Edelsberg J, Colditz GA, et al. Lifetime health and economic consequences of obesity. *Arch Intern Med* 1999; 159: 2177–2183.
4. Thompson D, Brown JB, Nichols GA, et al. Body mass index and future healthcare costs: a retrospective study. *Obes Res* 2001; 9: 210–218.
5. Gallagher SM. Meeting the needs of the obese patient. *Am J Nurs* 1996; 91: S1–S12.
6. Nowbar S, Bukart KM, Gonzales R, et al. Obesity-associated hypoventilation in hospitalized patients: prevalence, effects, and outcomes. *Am J Med* 2004; 116: 1–7.
7. Roth JL, Kushner RF, Bateman E. Nutritional support of the obese patient. *Curr Gastroenterol Rep* 2005; 7: 329–335.
8. Sharma AM, Pischon T, Engeli S, Scholze J. Choice of drug treatment for obesity-related hypertension: where is the evidence? *J Hypertens* 2001; 19: 667–674.
9. Sandernick GU, Le Liboux A, Jariwala, et al. The pharmacokinetics and pharmacodynamics of enoxaparin in obese volunteers. *Clin Pharmacol Ther* 2002; 72: 308–318.
10. Yee JY, Duffull SB. The effect of body weight on dalteparin pharmacokinetics: a preliminary study. *Eur J Clin Pharmacol* 2000; 56: 293–297.
11. Foley K, Lee RB. Surgical complications of obese patients with endometrial cancer. *Gynecol Oncol* 1990; 39: 171–174.
12. Braden BJ, Bergstrom N. Clinical utility of the Braden scale for predicting pressure sore risk. *Decubitus* 1989; 2: 44–51.
13. Mathison CJ. Skin and wound care challenges in the hospitalized morbidly obese patient. *J Wound Ostomy Continence Nurs* 2003; 30: 78–83.
14. Terry M, O'Brien SP, Kerstein MD. Lower-extremity edema: evaluation and diagnosis. *Wounds* 1998; 10: 118.
15. Hahler B. Morbid obesity: a nursing care challenge. *Dermatol Nurs* 2002; 14: 249–256.

16. Bergstrom N, Bennett MA, Carlson CE, et al. *Treatment of Pressure Ulcers*. Rockville, MD: US Department of Health and Human Services; 1994.; Clinical Practice Guideline 15; AHCPR Publication 95–0652.
17. Gallagher SM. Restructuring the therapeutic environment to promote care and safety for the obese patient. *J Wound Ostomy Continence Nurs* 1999; 26: 292–297.
18. Gallagher SM. Caring for the obese patient. *Nursing* 1998; 28: 32hn1–32hn3.
19. Gallagher SM. Meeting the needs of the obese patient. *Am J Nurs* 1996; 96: S1–S12.
20. Gallagher S, Gates J. Challenges of ostomy care and obesity. *Ostomy Wound Manage* 2004; 50: 38–40.
21. Colwell JC, Fichera A. Care of the obese patient with an ostomy. *J Wound Ostomy Continence Nurs* 2005; 32: 378–383.
22. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *JAMA* 2003; 289: 2560–2572.
23. Gress TW, Nieto FJ, Shahar E, et al. Hypertension and antihypertensive therapy as risk factors for type 2 diabetes mellitus. Atherosclerosis Risk in Communities Study. *N Engl J Med* 2000; 342: 905–912.
24. Messerli FH, Nunez BD, Ventura HO, Snyder DW. Overweight and sudden death. Increased ventricular ectopy in cardiopathy of obesity. *Arch Intern Med* 1987; 147: 1725–1728.
25. Grassi G, Cattaneo BM, Seravalle G, et al. Obesity and the sympathetic nervous system. *Blood Press* 1996; 1(suppl): 43–46.
26. Samuelsson O, Hedner T, Berglund G, et al. Diabetes mellitus in treated hypertension: incidence, predictive factors and the impact of non-selective beta-blockers and thiazide diuretics during 15 years treatment of middle-aged hypertensive men in the Primary Prevention Trial Goteborg, Sweden. *J Hum Hypertens* 1994; 8: 257–263.
27. De Courten M, Ferrari P, Schneider M, et al. Lack of effect of long-term amlodipine on insulin sensitivity and plasma insulin in obese patients with essential hypertension. *Eur J Clin Pharmacol* 1993; 44: 457–462.
28. Yusuf S, Sleight P, Pogue J, et al. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. *N Engl J Med* 2000; 342: 145–153.
29. Hansson L, Lindholm LH, Niskanen L, et al. Effect of angiotensin-converting-enzyme inhibition compared with conventional therapy on cardiovascular morbidity and mortality in hypertension: the Captopril Prevention Project (CAPPP) randomised trial. *Lancet* 1999; 353: 611–616.
30. Moan A, Risanger T, Eide I, Kjeldsen SE. The effect of angiotensin II receptor blockade on insulin sensitivity and sympathetic nervous system activity in primary hypertension. *Blood Press* 1994; 3: 185–188.
31. Engeli S, Negrel R, Sharma AM. Physiology and pathophysiology of the adipose tissue renin-angiotensin system. *Hypertension* 2000; 35: 1270–1277.
32. Sharma AM, Pischon T, Engeli S, Scholze, J. Choice of drug treatment for obesity-related hypertension: where is the evidence? *J Hypertens* 2000; 19: 667–674.
33. Craig WA. Pharmacokinetic/pharmacodynamic parameters: rationale for antibacterial dosing of mice and men. *Clin Infect Dis* 1998; 26: 1–10.
34. Pai MP, Bearden DT. Antimicrobial dosing considerations in obese adult patients. *Pharmacotherapy* 2007; 27: 1081–1091.
35. Wurtz R, Itokazu G, Rodvold K. Antimicrobial dosing in obese patients. *Clin Infect Dis* 1997; 25: 112–118.
36. Bearden DT, Rodvold KA. Dose adjustments for antibacterials in obese patients: applying clinical pharmacokinetics. *Clin Pharmacokinet* 2000; 38: 415–426.
37. Geerts WH, Heit JA, Clagett GP, et al. Prevention of venous thromboembolism: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest* 2004; 126(suppl): 338S–400S.
38. Stein PD, Beemath A, Olson RE. Obesity as a risk factor in venous thromboembolism. *Am J Med* 2005; 118: 978–980.
39. Heizmann M, Baerlocher GM, Steinmann F, et al. Anti-Xa activity in obese patients after double standard dose of nadroparin for prophylaxis. *Thromb Res* 2002; 106: 179–181.
40. Hamad GG, Choban PS. Enoxaparin for thromboprophylaxis in morbidly obese patients undergoing bariatric surgery: findings of the prophylaxis against VTE outcomes in bariatric surgery patients receiving enoxaparin (PROBE) study. *Obes Surg* 2005; 15: 1368–1374.
41. Kroon C, De Boer A, Kroon JM, et al. Influence of skinfold thickness on heparin absorption. *Lancet* 1991; 337: 945–946.
42. Nicolaides AN, Harbourne T, Russell RCG. Subtaneous heparin, plasma heparin levels and postoperative wound haematoma. *Br J Surg* 1984; 71: 62–64.
43. Shepherd MF, Rosborough TK, Schwartz ML. Heparin thromboprophylaxis in gastric bypass surgery. *Obes Surg* 2003; 13: 249–253.
44. Balant L. Clinical pharmacokinetics of sulfonylurea hypoglycaemic drugs. *Clin Pharmacokinet* 1981; 6: 215–241.
45. Jaber LA, Antal EJ, Slaughter RL, Welshman IR. The pharmacokinetics and pharmacodynamics of 12 weeks of glyburide therapy in obese diabetics. *Eur J Clin Pharmacol* 1993; 45: 459–463.
46. Jaber JA, Ducharme MP, Halapy H. The effects of obesity on the pharmacokinetics and pharmacodynamics of glipizide in patients with non-insulin-dependent diabetes mellitus. *Ther Drug Monitor* 1996; 18: 6–13.
47. Shukla UA, Chi EM, Lehr KH. Glimepiride pharmacokinetics in obese versus non-obese diabetic patients. *Ann Pharmacokinet* 2004; 38: 30–35.
48. Holmes G, Galitz L, Hu P, Lyness W. Pharmacokinetics of insulin aspart in obesity, renal impairment, or hepatic impairment. *Br J Pharmacol* 2005; 60: 469–476.
49. Padilla LA, Radosevich DM, Milad MP. Limitations of the pelvic examination for evaluation of the female pelvic organs. *Int J Gynecol Obstet* 2005; 88: 84–88.
50. Uppot RN, Sahani DV, Hahn PF, et al. Effect of obesity on image quality: fifteen-year longitudinal study for evaluation of dictated radiology reports. *Radiology* 2006; 240: 435–439.
51. Uppot RN. Impact of obesity on medical imaging. *Radiol Clin North Am* 2007; 188: 433–440.
52. Bae KT, Tao C, Gürel S, et al. Effect of weight and scanning duration on contrast enhancement during pulmonary multidetector CT angiography. *Radiology* 2007; 242: 582–589.

53. Burgstahler C, Beck TA, Reimann A, et al. Image quality and diagnostic accuracy of 16-slice multidetector computed tomography for the detection of coronary artery disease in obese patients. *Int J Obes* 2005; 30: 569–573.
54. Kussmaul WG, Bowers B, Dairywala I. Method for coronary angiography in morbidly obese patients. *Catheter Cardiovasc Interv* 2005; 65: 268–270.
55. McNulty PH, Ettinger SM, Field JM, et al. Cardiac catheterization in morbidly obese patients. *Catheter Cardiovasc Interv* 2002; 56: 174–177.
56. Unterborn J. Pulmonary function testing in obesity, pregnancy, and extremes of body habitus. *Clin Chest Med* 2001; 22: 759–767.
57. Crapo RO, Kelly TM, Elliott CG, Jones SB. Spirometry as a preoperative screening test in morbidly obese patients. *Surgery* 1986; 99: 763–768.
58. Plumb TJ, Adelson AB, Groggel GC, et al. Obesity and hemodialysis vascular access failure. *Am J Kidney Dis* 2007; 50: 450–454.
59. Vassalotti JA, Falk A, Cohl ED, et al. Obese and non-obese hemodialysis patients have a similar prevalence of functioning arteriovenous fistula using pre-operative vein mapping. *Clin Nephrol* 2002; 58: 211–214.
60. Choban PS, Flancbaum L. Nourishing the obese patient. *Clin Nutr* 2000; 19: 305–311.
61. Choban PS, Burge JC, Scales D, Flancbaum L. Hypoenergetic nutrition support in hospitalized obese patients: a simplified method for clinical application. *Am J Clin Nutr* 1997; 66: 546–550.
62. Roth JL, Kushner RF, Bateman E. Nutritional support of the obese patient. *Curr Gastroenterol Rep* 2005; 7: 329–335.
63. Shikora SA, Jensen GL. Hypoenergetic nutrition support in hospitalized obese patients. *Am J Clin Nutr* 1997; 66: 679–680.
64. Zizza C, Herring AH, Stevens J, Popkin BM. Length of hospital stay among obese individuals. *Am J Public Health* 2004; 94: 1587–1591.
65. de Pablo P, Losina E, Phillips CB, et al. Determinants of discharge destination following elective total hip replacement. *Arthritis Rheum* 2004; 51: 1009–1017.
66. Vincent HK, Weng JP, Vincent KR. Effect of obesity on inpatient rehabilitation outcomes after total hip arthroplasty. *Obesity* 2007; 15: 522–530.
67. Vincent HK, Vincent KR, Lee LW, Alfano AP. Effect of obesity on inpatient outcomes following total knee arthroplasty. *Clin Rehabil* 2007; 21: 182–190.
68. Yaegashi M, Jean R, Zurigat M, et al. *J Intensive Care Med* 2005; 20: 147–154.
69. Taggart HM, Mincer AB, Thompson AW. Caring for the orthopaedic patient who is obese. *Orthop Nurs* 2004; 23: 204–210.

Do Not Resuscitate Does Not Mean Do Not Treat: How Palliative Care and Other Modalities Can Help Facilitate Communication About Goals of Care in Advanced Illness

Cardinale B. Smith, MD[1,2] and Lynn Bunch O'Neill, MD[1]

[1]Hertzberg Palliative Care Institute, Brookdale Department of Geriatrics and Adult Development, Mount Sinai Medical Center, New York, NY
[2]Division of Hematology/Oncology, Mount Sinai Medical Center, New York, NY

ABSTRACT

A do-not-resuscitate (DNR) order is commonly used for hospitalized patients with advanced illness. It reflects only the desires of a patient once he or she suffers a full cardiopulmonary arrest. It does not reflect preferences about other forms of life-sustaining treatments. This article reviews the definition of a DNR order, describes the evidence suggesting that physicians use a DNR order in making determinations about other life-sustaining treatments, and will provide tools such as the use of palliative care consultations and combined directives in discussing overall goals of care. ***Mt Sinai J Med 75:460–465, 2008.*** *© 2008 Mount Sinai School of Medicine*

Key Words: Do-not-rescuscitate order, goals of care, palliative care, life-sustaining treatments, advanced directives.

Address Correspondence to:

Cardinale B. Smith
Hertzberg Palliative Care Institute
Brookdale Department of Geriatrics and Adult Development
Mount Sinai Medical Center
New York, NY
Email: cardinale.smith@mssm.edu

> I think the best physician is the one who has the providence to tell to the patients according to his knowledge the present situation, what has happened before, and what is going to happen in the future.
>
> Hippocrates

CLINICAL PROBLEM

A 78-year-old male is brought to the emergency room by his family with his second aspiration pneumonia event in 4 months. On admission, he is found to be febrile, tachycardic, and hypotensive. In addition, he has an altered mental status. The family states that he does have hypertension, congestive heart failure, and moderate dementia but normally is able to recognize them and communicate with them. He requires some assistance with his activities of daily living. He lives with his 75-year-old wife, and they have a home health aide 8 hours per day. The family states that this presentation is very similar to his prior hospitalization. At that time, he was treated with antibiotics and intravenous fluids and was able to return home after 7 days. His daughter believes that this is what his desire would be now. After a discussion with the physician, a do-not-resuscitate (DNR) order is placed by his daughter, who is also his healthcare proxy. The physician then tells the nurse, "This patient is comfort care now. Let's discontinue all nonessential medications, antibiotics, and intravenous fluids and limit blood draws." Is it correct to use a DNR order as a surrogate marker for the goals of care of patients and families rather than as a specific treatment preference?

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20076

© 2008 Mount Sinai School of Medicine

KEY ISSUES

The rate of long-term survival after in-hospital cardiac arrest, as reported in the medical literature, varies from 6.5% to 32% and on average is 15%.[1] For elderly patients with other comorbidities, the rate of long-term survival after cardiac arrest is probably no better than 5%.[2] Despite these facts, it is not surprising that patients and their families have an unrealistic view of what resuscitation can accomplish when contemporary media often portray the success rate as high as 70%.[2] Even when one adjusts for severity of illness, diagnosis, and other factors that might affect a critically ill patient's propensity to have a DNR order, the majority are likely to die in the hospital.[3–5]

A DNR order indicates that if a patient sustains a full cardiopulmonary arrest, no resuscitative measures are to be undertaken. In fact, the American Medical Association states that DNR orders preclude resuscitative efforts only in the event of cardiopulmonary arrest and should not influence other therapeutic interventions that may be appropriate for the patient.[6]

A DNR order may not preclude a patient from other life-sustaining treatments such as vasopressors, antibiotics, artificial nutrition and hydration, and intubation in the event of a purely respiratory event.

Despite this, there is well-recognized evidence, some from over a decade ago, that the plan of care for a patient with a DNR order is often unclear and fails to define limits other than cardiopulmonary resuscitation.[7–12] Data have also demonstrated that DNR orders are often broadly applied to other therapies.[7–12] One cross-sectional study conducted at 2 urban tertiary-care teaching hospitals examined the use of concurrent care concerns (CCCs), 11 previously validated indicators of the comprehensiveness of the plan of care.[7] These indicators included the use of intubation, dialysis, blood products, antibiotics, pressors, and artificial hydration and nutrition; analgesic and sedative needs; consideration of hospice; consideration of spiritual needs; and consideration of a decrease in the frequency of vital signs. The study found a mean of 1.55 CCCs per DNR order. In addition, it found independent associations between CCCs and severity of illness, with those having a greater severity of illness having more documented CCCs. Interestingly, 1 of the centers had educated its medical residents about patient care plans for patients with DNR orders 3 years prior to the start of the study, and the staff at this center attended to significantly more CCCs than the other center with no previous education on the topic. Another study surveyed primary and cross-covering residents of patients with DNR orders and found that residents intended to withhold a variety of other therapeutic interventions, that in half of these instances there was no chart documentation to that effect, and that there was little agreement between primary and cross-covering residents regarding which therapies to withhold.[8] Several retrospective chart reviews have found that patients in hospitals and long-term-care facilities with DNR orders receive fewer life-prolonging treatments, even after controlling for age, functional status, and severity of illness.[10,11] A cross-sectional survey mailed to medical residents and attending physicians, consisting of 3 patient scenarios followed by 10 treatment decisions that included or did not include a DNR order, found that physicians significantly agreed or strongly agreed to initiate fewer interventions when a DNR order was present versus when a DNR order was absent. In addition, patients with a DNR order were less likely to be transferred to an intensive care unit or to be intubated, and their physicians had a decreased willingness to draw blood cultures, perform central line placement, or give blood transfusions.[12]

Evidence suggests that physicians do not discuss goals of care as they pertain to DNR status because they have inadequate communication skills and because their preferences for patients are not in concordance with what patients actually want for themselves.[10,12,13] Furthermore, when the discussions are held, differences in the perceived outcome of the discussion may exist.[13–19] In 1 study, resident physicians, their patients, and family members were interviewed following a discussion regarding preferences for cardiopulmonary resuscitation, and it was found that in 6 of 28 interviews, the participants reported differing results of the discussion. These differences included 2 patients who desired resuscitation but had DNR orders placed and 2 patients who did not want to be resuscitated and did not have a DNR order placed.[19] This raises concern about the quality of communication with seriously ill hospitalized patients and their families. Addressing the goals of care for patients can increase the quality of care and improve communication between physicians, patients, and their families.

ARE THERE STRATEGIES TO IMPROVE THE PLAN OF CARE?

The Role of Palliative Care Consultation

Palliative care focuses on improving the quality of life of patients and their families who are facing the problems associated with chronic or terminal illness. It involves a whole patient rather than a disease-specific approach to patient care by addressing the physical, psychosocial, and spiritual aspects of illness. It can be provided at any time during a person's illness and can be delivered at the same time as curative and life-prolonging treatments. Palliative care aims to improve quality of life and relieve suffering by using specific skill sets, most notably by fostering continued communication between patients and physicians.

Establishing clear goals about the plan of care is the best way to facilitate the appropriate medical treatment in concordance with patient preferences. There are established guidelines for addressing goals of care (see Table 1).[20,21] Physicians can help facilitate discussions about goals of care by using open-ended questions. Examples include the following:

"What makes life worth living for you?"
"What are your most important hopes?"
"What are your biggest fears?"
"What would you consider to be a fate worse than death?"[21]

It is important to keep questions straightforward, listen carefully to the answers, and then follow the patient's lead by asking focused follow-up questions. Try to avoid language with unintended consequences, such as

"I think we should start pulling back."
"Do you want us to do everything possible?"
"There's nothing more we can do for you."[20]

Instead, try using language such as

"We will concentrate on improving your quality of life and minimizing your symptoms."
"I want to give you the best care possible."
"Let's discuss what we can do to meet your goal of ...."

Once the goals of care are established, it becomes much easier to construct a plan of care centered on those preferences.

Combined Advance Directives

Most advance directives of seriously ill patients do not guide medical decision-making beyond naming a healthcare proxy or documenting general preferences in a standard living will format. Even

Table 1. *Protocol To Negotiate Goals of Care.*

Recommendation	Comments
Create the proper setting.	• Before the meeting, determine the most appropriate participants (family members and other healthcare providers). • Allow adequate time. • Determine what to say before the meeting.
Clarify what the patient and family already know.	• "What have you been told about your medical situation so far?" • This allows you to correct any misinformation and tailor the conversation on the basis of their previous knowledge.
Explore the hopes and expectations of the patient and family.	• This allows you to distinguish between attainable and unattainable goals.
Suggest realistic goals.	• Suggest attainable goals based on the present clinical scenario and how they can best be achieved. • Review the appropriateness of disease-modifying treatments. • Try to explain, using simple language, why unrealistic goals cannot be met.
Use empathic responses.	• It is very important to allow silence and to listen. • Let the patient and family express emotions. • Once emotions are expressed, use a connector such as "I can see how upsetting this is to you."
Make a plan and follow through.	• Summarize the plan to ensure that your interpretation of the conversation and decisions is in agreement with the patient and family and indicate how the plan of care will meet their goals. • Make a plan for continued follow-up. • Inform the patient and family how to contact you if they have further questions or concerns. • Continue to review and revise the plan as needed.

NOTE: This table was adapted from the Education on Palliative and End-of-Life Care Project[20] and Buckman *et al.*[21]

when specific instructions are present, the execution of these preferences may be inconsistent.[22] Several intervention studies in which patients were asked to express choices of treatment that were then communicated to their physicians have documented success at eliciting and recording patient preferences. However, none have shown a difference in the use of life-sustaining treatment or in other desirable outcomes such as decreasing pain and suffering.[23-25] One such intervention study found that treatments were consistent with patient preferences 75% of the time, but the presence of an advance directive in the medical record did not necessarily result in a treatment that was more consistent with preferences.[23] When it was inconsistent with a patient's preference, treatment was usually less aggressive than that requested. Another intervention study demonstrated that offering advance directives to patients made no difference in the use of medical treatments, survival, health status, satisfaction, or the cost of care in the final month of life.[24] The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT) trial, a comprehensive intervention in which study nurses communicated detailed information on prognosis and preferences to physicians and their critically ill patients, had no effect on clinical decisions to use or to withhold life-sustaining treatments, pain control, or cost of terminal care.[25] Despite the clear success in recording patient treatment preferences near the end of life, these interventions could not alter other clinically desirable outcomes.[26] In a comprehensive physician education intervention that consisted of didactic sessions which were supplemented by chart reminders, group feedback, and the use of a new advance directive form in all inpatient charts, the number of recorded patient treatment preferences increased from 25% to 63%.[27] The researchers also measured a change in "impact directives," which were defined as preferences that would change treatment orders in the event of life-threatening complications, from 3% to 13%. This intervention was successful in broadening the range of treatments discussed.

Given this evidence, instructional or procedure-specific advance directives that more adequately address the complex issues that may arise with a critically or chronically ill individual have been suggested. These newer advance directives, called combined directives, merge components of instructional advance directives, the healthcare proxy, and the goals of care in an attempt to obtain a more complete understanding of patients' wishes.[28] The goal of these documents is to offer a more comprehensive advance care plan to guide clinicians in providing end-of-life care. These methods have different formats, with some being computer-based[29] and others paper-based.[30,31] Examples include "The Five Wishes Document from Aging with Dignity,"[30] which is legally valid in 40 states. This tool addresses the personal, emotional, spiritual, and medical needs of the patient and asks the following of the patient:

- Which person do you want to make healthcare decisions for you when you cannot make them?
- What kind of medical treatment do you want or not want?
- How comfortable do you want to be?
- How do you want people to treat you?
- What do you want your loved ones to know?

The Physician Orders for Life-Sustaining Treatment Paradigm (POLST) form,[31] which began in Oregon and is currently used in many states, is designed to express the individual's preferences for levels of treatment and either can indicate full treatment including resuscitation attempts or can be used to limit life-sustaining interventions. The form is most appropriate to use when the clinician feels that it would not be a surprise if the patient died in the next year. Various computer-based programs are largely institution-specific and are not uniform, but they provide an algorithm for physicians placing a DNR order for a hospitalized patient, directing the clinician to address not only a single DNR choice but also other plan-of-care options.

CONCLUSIONS AND RECOMMENDATIONS

Patients with DNR orders are often quite ill, and physicians may inappropriately assume that because these patients have already chosen to have 1 life-sustaining treatment withheld, they would also wish to have other life-sustaining treatments withheld. Yet patients may reasonably wish to avoid only cardiopulmonary resuscitation but still be interested in pursuing a variety of other therapies. Physicians should be educated in the communication skills necessary to undertake these discussions as they have been shown to lead to an increase in the use of patient preferences.[27] When possible and available, the use of a palliative care consultative team can also help in engaging in the discussions. There may also be benefit to adopting the use of combined directives to allow patient preferences to be documented. However, large-scale studies of combined directives have not yet been performed, and the relation of these directives to improving the quality of care is still largely unknown. Combined directives may be useful to the extent that they are able to prompt further dialogue with patients regarding goals of care.

SOLVING THE CLINICAL PROBLEM

In this case, it appears that the physician has misinterpreted the DNR order as a surrogate decision marker for other life-sustaining treatments. The patient described in the vignette, who has multiple chronic and serious medical conditions but is not actively dying, would benefit from the services of a palliative care consult. A discussion about realistic goals should be held with either the patient or his designated agent, with attention given to ascertaining the patient's opinion about an acceptable quality of life and establishing realistic attainable goals. This discussion can be summarized in a combined directive such as POLST or the "Five Wishes." Once goals are clarified, appropriate treatment can begin. It is possible that the patient could recover from this acute event, much like he recovered from the previous episode, and return to a baseline function that would be deemed to be a good quality of life by him and his family.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Zoch TW, Desbiens NA, DeStefano F, et al. Short- and long-term survival after cardiopulmonary resuscitation. *Arch Intern Med* 2000; 160: 1969–1973.
2. Diem SJ, Lantos JD, Tulsky JA. Cardiopulmonary resuscitation on television: miracles and misinformation. *N Engl J Med* 1996; 334: 1578–1582.
3. Shepardson LB, Youngner SJ, Speroff T, Rosenthal GE. Increased risk of death in patients with do-not-resuscitate orders. *Med Care* 1999; 37: 727–737.
4. Wenger NS, Pearson ML, Desmond KA, et al. Outcomes of patients with do-not-resuscitate orders: towards an understanding of what do-not resuscitate orders mean and how they affect patients. *Arch Intern Med* 1995; 155: 2063–2068.
5. DeJonge KE, Sulmasy DP, Gold KG, et al. The timing of do-not-resuscitate orders and hospital costs. *J Gen Intern Med* 1999; 14: 190–192.
6. American Medical Association, Council on Ethical and Judicial Affairs. "Guidelines for the Appropriate Use of Do-Not-Resuscitate Order." Journal of the American Medical Association 265 (1991):1,868-1,871.
7. Sulmasy DP, Sood JR, Ury, WA. The quality of care plans for patients with do-not-resuscitate orders. *Arch Intern Med* 2004; 164: 1573–1578.
8. Uhlmann RF, Cassel CK, McDonald WJ. Some treatment-withholding implications of no-code orders in an academic hospital. *Crit Care Med* 1984; 12: 879–881.
9. La Puma J, Silverstein MD, Stocking CB, et al. Life-sustaining treatment. A prospective study of patients with DNR orders in a teaching hospital. *Arch Intern Med* 1988; 148: 2193–2198.
10. Bedell SE, Pelle D, Maher PL, et al. Do-not-resuscitate orders for critically ill patients in the hospital. How are they used and what is their impact? *JAMA* 1986; 256: 233–237.
11. Holtzmann J, Phelay AM, Lurie N. Changes in orders limiting care and the use of less aggressive care in a nursing home population. *J Am Geriatr Soc* 1994; 42: 275–279.
12. Beach MC, Morrison RS. The effect of do-not-resuscitate orders on physician decision-making. *J Am Geriatr Soc* 2002; 50: 2057–2061.
13. Haidet P, Hamel MB, Davis RB, et al. Outcomes, preferences for resuscitation and physician-patient communication among patients with metastatic colon cancer. SUPPORT Investigators. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments. *Am J Med* 1998; 105: 222–229.
14. Smith AK, Ries AP, Zhang B, et al. Resident approaches to advance care planning on the day of hospital admission. *Arch Intern Med* 2006; 166: 1597–1602.
15. Tulsky JA, Chesney MA, Lo B. See one, do one, teach one? *Arch Intern Med* 1996; 156: 1285–1289.
16. Tulsky JA, Chesney MA, Lo B. How do medical residents discuss resuscitation with patients? *J Gen Intern Med* 1995; 10: 436–442.
17. Fischer GS, Tulsky JA, Rose MR, Arnold RM. Opening the black box: how do physicians communicate about advance directives? *Ann Intern Med* 1998; 129: 441–449.
18. Teno JM, Hakim RB, Knaus WA, et al. Preferences for cardiopulmonary resuscitation: physician-patient agreement and hospital resource use. SUPPORT Investigators. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments. *J Gen Intern Med* 1995; 10: 179–186.
19. Deep KS, Grif?th CH, Wilson JF. Discussing preferences for cardiopulmonary resuscitation: what do resident physicians and their hospitalized patients think was decided? *Patient Educ Couns* 2008; 72: 20–25.
20. The EPEC Project: Education on Palliative and End-of-life Care. Available at: http://www.epec.net. Accessed March 2008.
21. Baile WF, Buckman R, Lenzi T, et al. SPIKES–a six-step protocol for delivering bad news. *Oncologist* 2000; 5: 302–311.
22. Teno JM, Licks S, Lynn J, et al. Do advance directives provide instructions that direct care? *J Am Geriatr Soc* 1997; 45: 508–512.
23. Danis M, Southerland LI, Garrett JM, et al. A prospective study of advance directives for life-sustaining care. *N Engl J Med* 1991; 324: 882–888.

24. Schneiderman LJ, Kronick R, Kaplan RM, et al. Effects of offering advance directives on medical treatments and costs. *Ann Intern Med* 1992; 117: 599–606.
25. A controlled trial to improve care for seriously ill hospitalized patients. The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT). The SUPPORT Principal Investigators. *JAMA* 1995; 274: 1591–1598.
26. Hanson LC, Tulsky JA, Danis M. Can clinical interventions change care at the end of life? *Ann Intern Med* 1997; 126: 381–388.
27. Reilly BM, Wagner M, Magnussen CR, et al. Promoting inpatient directives about life-sustaining treatments in a community hospital. Results of a 3-year time-series intervention trial. *Arch Intern Med* 1995; 155: 2317–2323.
28. Collins LG, Parks SM, Winter L. The state of advance care planning: one decade after SUPPORT. *Am J Hosp Palliat Care* 2006; 23: 378.
29. Heffner JE, Barbieri C, Fracica P, Brown LK. Communicating do-not-resuscitate orders with a computer-based system. *Arch Intern Med* 1998; 158: 1090–1095.
30. Aging with Dignity. The Five Wishes. Available at: http://www.agingwithdignity.org/5wishes.html. Accessed March 2008.
31. Tolle SW, Tilden VP, Nelson CA, Dunn PM. A prospective study of the ef?cacy of the physician order form for life-sustaining treatment. *J Am Geriatr Soc* 1998; 46: 1097–1102.

ETHICS

From "I'm Not Staying!" to "I'm Not Leaving!": Ethics, Communication, and Empathy in Complicated Medical Discharges

David Alfandre, MD

Division of General Internal Medicine, Mount Sinai School of Medicine, New York, NY

Problems with hospital discharges present unique ethical dilemmas that physicians often find challenging to manage. From patients that leave against medical advice to those that refuse to leave, understanding the ethical dimensions of these dilemmas can help to improve the quality of care that patients receive. Once the ethical dimensions are clarified and understood, however, the success of the intervention is dependent on the quality of the communication between the patient and doctor. Although deficits in communication are at the root of many ethical dilemmas,[1] the skillful use of evidenced-based communication strategies can help in resolving others.[2,3] This article explores the problem of difficult hospital discharges by examining how these issues interact. Using a set of clinical cases, the article provides an evidenced-based introduction to the dynamic interplay of the ethical, legal, institutional, and interpersonal issues involved in discharge dilemmas.

SIGNING OUT AGAINST MEDICAL ADVICE

Ms. TH is a 67-year-old African American woman with a history of hypertension and peripheral vascular disease who has been admitted to the housestaff service with a diagnosis of hypertensive urgency presumed to be due to medication nonadherence. In the emergency room, she was asymptomatic with a systolic blood pressure of 220/114 and had a normal physical examination and normal laboratory work. On the day following admission, her systolic blood pressure remains over 190, and she tells the resident that she wishes to leave the hospital. When the resident spends the next 35 minutes discussing the risks and benefits of her decision and alternatives, she angrily insists that she is "worn out" and wants to go home because she does not like staying in the hospital. The resident is concerned that this is a "bad" decision that is not in her best interest and so turns to the hospital ethics consultant for advisement.

Patients who choose to leave the hospital prematurely and against the advice of their physician represent a common and challenging dilemma. In studies examining the problem, the discharged against medical advice rate was 1% to 2% of all admissions,[4,5] with a calculated readmission cost 50% higher than the original admission cost.[6] These patients who sign out against medical advice also have a higher rate of readmission.[7] Asthma patients discharged against medical advice have a 30-day 4 times greater risk of readmission to the emergency room and a 3 times greater risk of readmission to the hospital.[8]

Discharge against medical advice causes a considerable amount of concern among the providers who care for these patients.[9,10] Physicians relying too heavily on a patient's self-determination without a careful and thorough evaluation of the patient's decision may be inadvertently abandoning the patient at a time of vulnerability.[11] Indeed, finding this

Address Correspondence to:

David Alfandre, MD
Internal Medicine Associates
New York, NY
Email: david.alfandre@
mountsinai.org

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20067

© 2008 Mount Sinai School of Medicine

balance between attempting to protect patients from making "bad" medical decisions and permitting patients to fully exercise their autonomy is a central conflict in many ethical dilemmas. This issue is particularly important because sick patients and their families may have difficulty understanding and integrating their doctor's risk-benefit discussions.[12–14] Framed in terms of ethical principles, this problem is characterized by the tension between beneficence for the patient and respect for the patient's autonomy.[15] In this particular case, it weighs doing what is best for the patient, that is, keeping her in the hospital to improve her blood pressure, against allowing this competent self-determining patient to leave prematurely at uncertain medical risk.

Framing this issue only in terms of ethical principles, although useful as an initial and general organizational structure for thinking critically about the problem, gives very little guidance on practically approaching the issue with patients.[16,17] Indeed, this reliance solely on bioethical principles fails to either offer the critical interpersonal skills needed to manage a solution or ensure that the concerns of all the stakeholders in the process are accommodated.[18]

As for the aforementioned case, the issue of "bad" medical decisions has been further characterized in the literature on motivational interviewing.[19] "Bad" is a poor and pejorative descriptor in these types of dilemmas because it inadvertently prioritizes health concerns over all others, and this is a decision that not all patients share. Longevity is only one of many values that patients routinely consider in making decisions about their health. Rather than evaluating decisions through a physician-centered health model, providers can use motivational interviewing to examine decisions through the perspective of the patient. Thus, healthcare decisions are evaluated by their impact on the patient's life through an exploration of the inherent ambivalence of the positive and negative aspects of both the status quo and change.[20,21]

Taken in the context of discharge decisions and in light of the fact that there remains a dearth of literature on the motivations for patients' decisions to sign out against medical advice, the use of this empathic approach can help providers to understand how patients make these decisions. Better information about a patient's motivating behavior allows clinicians to target their counseling more appropriately and potentially negotiate a discharge to a more medically appropriate time. For example, when a clinician determines that a patient wants to leave the hospital because she feels pressure to return promptly to work, he can attempt to reduce that burden by focusing on that issue (perhaps advocating on the patient's behalf to her employer) rather than the perceived discharge conflict between the doctor and patient.

Following the evidence-based literature on informed consent can maximize the possibility that patients who choose to sign out against medical advice are fully evaluated. An informed decision means that the patient has arrived at her decision by consultation with her provider, not under duress or coercion, and with an understanding of the risks and benefits of the issue at hand and alternatives. Part of this process requires a formal capacity evaluation. Although most patients are not evaluated for capacity prior to a discharge against medical advice,[22,23] performing this important evaluation at the bedside ensures that a patient is making a fully informed decision and should reassure the physician that the patient has received complete and appropriate care.

A formal decisional capacity evaluation includes the following:

> Does the patient understand and appreciate the diagnosis, prognosis, likelihood of risks and benefits, and treatment alternatives?
> Can the patient make and communicate a choice?
> Can the patient articulate a reason for the refusal that is consistent with her values?[24]

Many advocate the use of a "sliding scale" of capacity assessment: the greater the risk is due to the patient's refusal, the more certain the clinician should be that the patient has decisional capacity.[25] For example, patients may have the capacity to refuse a blood draw but not to refuse a life-saving blood transfusion.

The resident and ethics consultant return together to the bedside to discuss her case further. When the consultant sits down, he introduces himself and asks the patient more about her illness. The patient responds that she knows her blood pressure is high and that if it stays high it could cause a stroke or other problem. Upon further questioning from the consultant, the patient reports that she currently has a disabled husband at home who needs her help cooking, cleaning, and bathing. She reports that she has been feeling overwhelmed lately with her husband's illness and was relieved to have had a "night off" while staying in the hospital. She agrees to take her medicine but would now like to return home to care for her husband. The patient agrees to have her medicine titrated up over the course of the morning, so that the team can normalize her blood pressure, and she can return home by late afternoon.

Many ethical dilemmas find their origin in incomplete communication between the provider and patient.[1] Furthermore, when emotions become heightened, as they did in this case, it is often challenging for providers to "maintain their center" and not respond directly to the patient's negative emotion. Indeed, mindfulness of their own emotions in stressful patient encounters can help providers to ensure that the empathic focus remains on the patient[26,27] to maintain the therapeutic alliance. As the patient became angry in response to the resident's repeated questions, there arose an opportunity to explore her history further. In this way, history taking often does not proceed in a linear fashion[28] but responds to the particular needs of the patient. Critical moments of the interview, so-called empathic opportunities, arise when the patient offers a clue about his underlying psychosocial state[2] (in this case, the patient reported feeling worn out). Attention to these clues ensures that the physician addresses informed consent and the range of needs of the patient and may help to avoid these types of dilemmas.[29]

PATIENTS WHO REFUSE TO LEAVE THE HOSPITAL

James S is a 48-year-old male paraplegic admitted with hematuria. His paraplegia is due to remote neurosyphilis that is complicated by multiple decubiti, urinary incontinence, and urethral strictures. He uses a manual wheelchair. He also has chronic pain and insomnia due to anxiety which are treated with high doses of narcotics and benzodiazepines, respectively. The admitting medical team promptly evaluates the chief concern of the patient and gathers consultations from the following services: urology to manage the hematuria, rehabilitation medicine to make recommendations for the paraplegia, psychiatry to assist with the insomnia and anxiety medications, and social work to assist in discharge planning.

Patients who refuse to leave the hospital are less common than patients who decide to sign out against medical advice. Both populations, however, are at risk for medical complications, the former being at risk for hospital-acquired infections. Discharges from a medical service, as illustrated in this case, should begin at the time of admission. In patients with complicated medical and psychiatric illnesses, without forethought and planning, discharge from the hospital can be delayed for days and sometimes weeks if insufficient resources are not gathered early in the course of the admission.[30] Among these, the input of specialty services is clearly vital in the management of patients with multiple comorbidities, but the primary team is in the unique position of advocating for and representing the patient if and when disagreement or conflict arises.

Mr. S is evaluated by the urology service. On the basis of its diagnostic workup, the urology service concludes that the patient has ureteral strictures that are the cause of his hematuria. Because he is clinically stable, the urology service recommends no emergent management but rather an outpatient treatment plan for potential ureteral stent placement.

Rehabilitation medicine recommends outpatient management as well for his paraplegia and sets up follow-up in its wheelchair clinic for the patient. The patient refuses to see a psychiatrist but offers no reason for his refusal. The primary team continues to prescribe him his outpatient doses of anxiolytics and hypnotics for his insomnia as well as narcotics for his chronic pain, although they are concerned about what they perceive as extraordinarily high doses of both. Because the medical workup has been completed, the primary team begins to finalize the patient's care plan in anticipation of discharge.

Ensuring a safe and appropriate discharge from the hospital is an ethical, legal, and institutional imperative and incorporates multiple processes of care.[31] A patient's primary medical problems must be evaluated and treated to a such degree that further management can take place safely on an outpatient basis. Furthermore, the essential psychosocial and economic considerations relevant to a patient's discharge must be addressed. Legal guidelines for the hospital to ensure a safe discharge are generally dictated by similar Medicare and state Medicaid policies[32] and reflect the aforementioned principles. A patient's request for further medical care, however, if not medically indicated, illustrates the ethical principle of justice, in this setting being manifest in the fair distribution of healthcare resources.

The patient reports that he would like to return home to stay with his sister who lives nearby. He reports living alone but can no longer afford that option. The social worker meets with the patient to gather the information to ensure that the discharge is appropriate and can be arranged with the sister. The patient reports that he does not have his sister's telephone number and asks the team to return when his sister calls him again during the evening. That night, he spends hours at the nursing station demanding more benzodiazepine medication to help him sleep. The following morning, when the intern arrives to discuss his discharge, the patient requests that the team "leave him alone" and come back in the afternoon. This happens repeatedly over the course of the following 24 hours.

The social worker meets with the team to report that on the basis of her discussion with the patient's sister, the patient has been homeless for the last 6 months, living in various shelters, and has severed ties with his sister. The sister is unable and unwilling to have the patient stay with her. When the team returns to the patient's bedside to discuss these findings, he becomes visibly angry and refuses to discuss the issue any further. The social worker returns later in the day to the patient to discuss shelter options that accommodate disabled individuals. When the patient refuses to return to a shelter, the social worker suggests that she investigate subacute rehabilitation facilities that can attend to the patient's multiple medical problems. He refuses those options as well.

Some parts of the preceding scenario can be avoided or at least mitigated by careful attention to communication strategies. In this scenario, there were multiple interactions when the house officers chose to avoid patient confrontation concerning both medical and psychiatric issues. Although this may be done for the sake of expediency and potentially because of the lack of physician training in this skill set,[33] the resulting situation is one that can turn fractious and result in barriers in the doctor-patient relationship.

Failing to collect the relevant social history early in the admission delayed "diagnosing" a key element in this patient's overall evaluation. When the patient's homelessness was discovered later in the admission, leading to delays in discharge, it left the primary team frustrated and interfered in an empathic evaluation.

When the primary team did not further inquire about the patient's past medical history and use of both benzodiazepines and narcotics, they were left with a major gap in the patient's medical and psychiatric history. Neglecting this part of the evaluation led to deficiencies in care as the team continued to prescribe medication that may have not been indicated. When the patient refused to be evaluated by a specialist for these problems, the team again missed an opportunity to further evaluate what may have been a medication abuse problem. Perhaps the patient's homelessness was due to his prescription dependence on narcotics? Although the ethical and professional imperative remained palliation of pain, the idea that his increasingly aggressive behavior was related to a substance abuse problem should have been evaluated. In an evaluation for drug abuse, urine toxicology has great utility[34–36] because the presence of illicit substances in the urine can indicate that the patient has a comorbid substance abuse disorder that would benefit from evaluation and treatment.

Because substance abuse is associated with discharge dilemmas, clearly and empathically approaching the situation can help to improve the quality and efficiency of care for this underserved population.[6] How an evaluation for drug abuse is discussed by the provider can affect the patient's response. Should a provider take an accusatory tone, the patient is likely to respond in kind. Maintaining a similar empathic attitude toward all patients, regardless of diagnosis, is likely to result in a clearer evaluation that reflects the core professional imperatives of being a physician.[37,38] An angry reaction by a patient to an empathic exploration of substance abuse may represent either something about the patient's potential pathology[39] or simply poorly treated pain that has been historically ignored by other providers. A clear nonjudgmental dialogue is critical to a thorough and objective evaluation.

The intern returns to the bedside. "I'd like to talk to you some more about your care. We believe that we have fully evaluated your concern about the blood in your urine. Our specialists, the computed tomography scan, and the blood work all indicate that the bleeding is not serious and that no further medical issues need to be attended to in the hospital. They should, however, be taken care of after you leave the hospital. A safe discharge plan is our next step and one that we need your help with. You have refused transfer to a shelter or a subacute rehabilitation facility, unfortunately our only options at this point. I'd like to hear your concerns so that we can come to an agreement that both of us can live with." The patient responds angrily, "You treat me like a drug addict. I'm not going to listen to this anymore." The intern responds, "I'm sorry you feel that way, and I wish that there were other options at this point. You'll have to decide among them. Let me know how I can help."

The patient refuses to discuss the issue further but decides to accept discharge to a subacute rehabilitation facility.

Reacting empathically and nonjudgmentally to angry patients allows the primary physician to model the professionalism that is so often more elusive when one is dealing with difficult-to-help patients. When the treating team does not escalate the conflict with the patient, it is easier to focus on the imperatives of care and not the conflict. Furthermore, when the physician permits the patient to report his concerns, a more targeted approach to a shared solution is possible.

Ethical dilemmas are certainly unavoidable in the regular care of patients. However, careful attention to potential deficits in communication and misunderstandings about patients' values and concerns can help to limit the frequency and complexity of these dilemmas. Using ethical principles and legal principles to understand and form a structure for understanding ethical dilemmas surrounding discharges is only part of the solution. Communication skills, empathic listening, and a focus on evidenced-based care can ensure that discharges are done safely, timely, and with the fully informed participation of the patient.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Swetz KM, Crowley ME, Hook C, Mueller PS. Report of 255 clinical ethics consultations and review of the literature. *Mayo Clin Proc* 2007; 82: 686–691.
2. Levinson W, Gorawara-Bhat R, Lamb J. A study of patient clues and physician responses in primary care and surgical settings. *JAMA* 2000; 284: 1021–1027.
3. Epstein RM, Hadee T, Carroll J, et al. "Could this be something serious?" Reassurance, uncertainty, and empathy in response to patients' expressions of worry. *J Gen Intern Med* 2007; 22: 1731–1739.
4. Ibrahim SA, Kwoh CK, Krishnan E. Factors associated with patients who leave acute-care hospitals against medical advice. *Am J Public Health* 2007; 97: 2204–2208.
5. Saitz R, Ghali WA, Moskowitz MA. The impact of leaving against medical advice on hospital resource utilization. *J Gen Intern Med* 2000; 15: 103–107.
6. Aliyu ZY. Discharge against medical advice: sociodemographic, clinical and financial perspectives. *Int J Clin Pract* 2002; 56: 325–327.
7. Hwang SW, Li J, Gupta R, et al. What happens to patients who leave hospital against medical advice? *CMAJ* 2003; 168: 417–420.
8. Baptist AP, Warrier I, Arora R, et al. Hospitalized patients with asthma who leave against medical advice: characteristics, reasons, and outcomes. *J Allergy Clin Immunol* 2007; 119: 924–929.
9. Taqueti VR. Leaving against medical advice. *N Engl J Med* 2007; 357: 213–215.
10. Carrese JA. Refusal of care: patients' well-being and physicians' ethical obligations: "but doctor, I want to go home". *JAMA* 2006; 296: 691–695.
11. Quill TE, Brody H. Physician recommendations and patient autonomy: finding a balance between physician power and patient choice. *Ann Intern Med* 1996; 125: 763–769.
12. Pochard F, Azoulay E, Chevret S, et al. Symptoms of anxiety and depression in family members of intensive care unit patients: ethical hypothesis regarding decision-making capacity. *Crit Care Med* 2001; 29: 1893–1897.
13. Sheehan J, Sherman KA, Lam T, Boyages J. Association of information satisfaction, psychological distress and monitoring coping style with post-decision regret following breast reconstruction. *Psychooncology* 2007; 16: 342–351.
14. Apkarian AV, Sosa Y, Krauss BR, et al. Chronic pain patients are impaired on an emotional decision-making task. *Pain* 2004; 108: 129–136.
15. Beauchamp TL, Childress JF. *Principles of Biomedical Ethics*. 5th ed. New York, NY: Oxford University Press; 2001.
16. Gillon R. Medical ethics: four principles plus attention to scope. *BMJ* 1994; 309: 184–188.
17. Fiester A. Viewpoint: why the clinical ethics we teach fails patients. *Acad Med* 2007; 82: 684–689.
18. Dubler N. The importance of interpersonal skills in multidisciplinary education. In: Kopelman LM, ed. Building Bioethics: Conversations with Clouser and Friends on Medical Ethics 1st ed. London, United Kingdom: Kluwer Academic Publishers; 1999:, 37–49.
19. Miller S, Rollnick W. *Motivational Interviewing–Preparing People for Change*. 2nd ed. New York, NY: Guildford Press; 2002.
20. Rubak S, Sandbaek A, Lauritzen T, Christensen B. Motivational interviewing: a systematic review and meta-analysis. *Br J Gen Pract* 2005; 55: 305–312.
21. Emmons KM, Rollnick S. Motivational interviewing in health care settings. Opportunities and limitations. *Am J Prev Med* 2001; 20: 68–74.
22. Seaborn Moyse H, Osmun WE. Discharges against medical advice: a community hospital's experience. *Can J Rural Med* 2004; 9: 148–153.
23. Henson VL, Vickery DS. Patient self discharge from the emergency department: who is at risk? *Emerg Med J* 2005; 22: 499–501.
24. Lo B. *Resolving Ethical Dilemmas: A Guide for Clinicians*. 2nd ed. New York, NY: Lippincott; 2005.
25. Drane JF. Competency to give an informed consent. A model for making clinical assessments. *JAMA* 1984; 252: 925–927.
26. Levinson W. Frustrating patients: using our feelings as diagnostic clues. *J Gen Intern Med* 1991; 6: 259–260.
27. Halpern J. Empathy and patient-physician conflicts. *J Gen Intern Med* 2007; 22: 696–700.
28. Haidet P, Paterniti DA. "Building" a history rather than "taking" one: a perspective on information sharing during the medical interview. *Arch Intern Med* 2003; 163: 1134–1140.
29. Epstein RM, Shields CG, Franks P, et al. Exploring and validating patient concerns: relation to prescribing for depression. *Ann Fam Med* 2007; 5: 21–28.
30. Yates FD Jr. Holding the hospital hostage. *Am J Bioeth* 2007; 7: 36–37.
31. Banja J, Eig J, Williams MV. Discharge dilemmas as system failures. *Am J Bioeth* 2007; 7: 29–31.
32. Swidler RN, Seastrum T, Shelton W. Difficult hospital inpatient discharge decisions: ethical, legal and clinical practice issues. *Am J Bioeth* 2007; 7: 23–28.
33. DuVal G, Clarridge B, Gensler G, Danis M. A national survey of U.S. internists' experiences with ethical dilemmas and ethics consultation. *J Gen Intern Med* 2004; 19: 251–258.
34. Manchikanti L, Manchukonda R, Pampati V, et al. Does random urine drug testing reduce illicit drug use in chronic pain patients receiving opioids? *Pain Physician* 2006; 9: 123–129.
35. Katz NP, Sherburne S, Beach M, et al. Behavioral monitoring and urine toxicology testing in patients receiving

long-term opioid therapy. *Anesth Analg* 2003; 97: 1097.

36. Katz N, Fanciullo GJ. Role of urine toxicology testing in the management of chronic opioid therapy. *Clin J Pain* 2002; 18(suppl): S76–S82.
37. Passik SD, Byers K, Kirsh KL. Empathy and the failure to treat pain. *Palliat Support Care* 2007; 5: 167–172.
38. Gallagher RM. Empathy: a timeless skill for the pain medicine toolbox. *Pain Med* 2006; 7: 213–214.
39. Lipkin M. The medical interview. In: Feldman M, Christensen J, eds. *Behavioral Medicine in Primary Care*. 1st ed. Stamford, CT: Appleton & Lange; 1997; 1–7.

Mount Sinai Journal of Medicine 75:472–477, 2008

EVIDENCE-BASED MEDICINE

Clinical Prediction Rules for Hospitalists

Ramiro Jervis, MD, and Thomas McGinn, MD

Mount Sinai School of Medicine, New York, NY

Hospitalists struggle like all physicians to incorporate the best evidence into daily practice. The busy work flow of the hospital, however, with an emphasis on both throughput and quality, produces unique pressures to diagnose and treat patients as efficiently as possible. These pressures highlight a need to bring the practice of evidence-based medicine (EBM) to the bedside of patient care in the most efficient manner possible.

Clinical prediction rules (CPRs) form a unique subset of the practice of EBM and strive to address this challenge of using EBM at the bedside. A CPR is a clinical tool that quantifies the individual contributions that various components of the history, physical examination, and basic laboratory results make toward the diagnosis, prognosis, or likely response to treatment of a patient. Clinical decision rules attempt to formally test, simplify, and increase the accuracy of clinicians' diagnostic and prognostic assessments.[1]

CPRs are most likely to be useful in situations in which decision making is complex, the clinical stakes are high, or there are opportunities to achieve cost savings without patient care being compromised.[1] Many examples of CPRs exist, and this article provides a sampling of clinical scenarios that highlight the utility of these tools for hospitalists, followed by a discussion on assessing the validity of CPRs before they are incorporated into daily practice.

Address Correspondence to:

Ramiro Jervis, MD
Mount Sinai School of Medicine
New York, NY
Email: ramiro.jervis@
mountsinai.org

CLINICAL PREDICTION RULES: CASE EXAMPLES

Using Clinical Prediction Rules To Assess Pretest Probability for Pulmonary Embolism

Pulmonary embolism (PE) remains a major public health problem and by some estimates accounts for 10% of all hospital deaths.[2] The importance of accurate and efficient diagnoses of thromboembolic disease makes this disease an ideal starting point for CPRs. Pretest probability, defined as the probability of disease before the test result is known, is often a challenging variable to less experienced clinicians and plays a key role in the diagnoses of many disease, including PE. CPRs can simplify and standardize the assessment of a patient's pretest probability of disease. Consider this typical case admitted by a hospitalist.

A 50-year-old woman with a history of orthopedic surgery 3 weeks prior and no past medical history presents with acute onset of pleuritic chest pain and shortness of breath while lifting a heavy object. Her pain lasted only seconds, but her shortness of breath persists. She has no leg swelling or pain. She admits to symptoms of anxiety and stress. Her physical examination demonstrates a blood pressure of 100/70, a pulse of 80, and an O_2 saturation of 97% on room air. Her chest X-ray is clear, her electrocardiogram is normal, and her laboratory work and arterial blood gas are unremarkable. The patient is admitted to the medical floor while her workup is pending.

The decision for further action for this patient is highly dependent on the clinician's perception of risk for PE, and the interpretation of further results is dictated by the pretest probability. If, for example, the patient in question is low-risk by clinical intuition, a low-probability ventilation perfusion (VQ) scan may be sufficient for a clinician to safely discharge the patient. Similarly, if the patient is perceived as

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20066

© 2008 Mount Sinai School of Medicine

medium- or high-risk, a low-probability VQ scan may not be sufficient, and further testing with D-dimer, lower extremity Dopplers, computed tomography (CT) angiography, or pulmonary angiography may be required.

This dependence on pretest probability is not restricted to testing with a VQ scan. CT pulmonary angiography, for example, has a negative likelihood ratio of only 0.2. (Likelihood ratios incorporate the sensitivity and specificity of a test and provide insight into how much the pretest probability goes up or down with a given test result. A further and more in-depth discussion of likelihood ratios can be found in Grimes and Schulz.[3]) A high pretest probability of 67%, for example, would be reduced only to a 12% probability of disease with a negative CT angiogram, which is unacceptable for many clinicians, given the morbidity and mortality associated with PE.[4] Although exact numbers like this can be helpful, most clinicians will intuitively understand that a high clinical suspicion requires a more robust burden of proof to counter a diagnosis, whereas a low clinical suspicion of disease needs fewer confirmatory negative tests. The diagnosis of PE and many other diseases for that matter is then highly dependent on an accurate pretest probability, regardless of the diagnostic modality.

CPRs may be especially useful for less experienced clinicians,[5] and one of the most widely used CPRs for predicting thromboembolism is the Simplified Wells Scoring System.[6] This scoring system (Table 1) looks at different aspects of the history and physical examination and weighs them to calculate a score for risk. The Wells scoring system (also known as the Wells criteria) has compared favorably with experienced clinical intuition, with disease probabilities for PE in 10 prospective studies ranging from 3% to 28%, from 16% to 46%, and from 38% to 98% in the low, moderate, and high pretest probability groups, respectively.[5] Defining the lower risk groups is helpful to the less experienced clinician, as fewer additional tests (such as D-dimer or lower extremity ultrasound) will be necessary for this group to reduce the disease probability to a more comfortable range prior to discharge.

Case Resolution

The patient has no signs or symptoms suggestive of DVT. An alternative diagnosis of musculoskeletal pain and anxiety is at least as likely as a diagnosis of PE. There is no tachycardia as defined under the Wells criteria, and there is no history of thromboembolism, cancer, or hemoptysis. Only a history of recent surgery gives the patient a total of 1.5 points on the Wells score; this is defined as low-risk because the score is less than 2. The doctor decides that the patient does not need a VQ scan or CT angiogram and that the patient can be safely discharged if a D-dimer is negative. Recent studies, in fact, validate this approach.[7] The patient is subsequently discharged with outpatient follow-up.

Predicting Risk for Pneumonia

Pneumonia is a common ailment treated by hospitalists and may be associated with high mortality in certain populations. Low-risk patients, however, can be treated in the community if they can be reliably identified. Such risk assessment then is crucial to the mission of hospitalists in reducing unnecessary resource utilization. Consider the following typical case.

A healthy 62-year-old patient is admitted to a service with a diagnosis of pneumonia. She has a history of controlled hypertension and diabetes and had presented to the Emergency Department with productive cough and fever at home. A left lower lobe infiltrate is present on chest X-ray. Her laboratory work is unremarkable except for a white count of 18,000. The patient is alert and oriented, has a temperature of 38.5°C, a blood pressure of 110/70, and an oxygen saturation of 95% on room air, and

Table 1. *Wells Clinical Prediction Rule for PE (Simplified Criteria).*

Points	If the Total Score Is	PE Probability Is
• Clinical symptoms of PE: chest pain, shortness of breath, and oxygen saturation <90% (3.0 points)	<2.0 points	Low
• No alternative diagnosis (3.0 points)	2.0–6.0 points	Moderate
• Heart rate >100 (1.5 points)	<6.0 points	High
• Immobilization or surgery in the previous 4 weeks (1.5 points)		
• Previous DVT/PE (1.5 points)		
• Hemoptysis (1 point)		
• Malignancy (1 point)		

Abbreviations: DVT, deep vein thrombosis; PE, pulmonary embolism.

is breathing at a rate of 20. The patient wants to go home but "only if the physicians think it is safe."

Much effort has gone into developing risk stratification tools to safely triage patients who may or may not need further inpatient management. The first of these tools by Fine *et al.*[8] was developed in 1997 and comprises 20 clinical variables to triage patients into 5 risk categories, with risk of death ranging from 0.5% for inpatients stratified to the lowest risk group to 27% 30-day mortality for the highest risk group. Although handheld programs or online calculators can assist in the calculation of these scores, the complexity of the Pneumonia Severity Index (PSI) may limit its everyday use. The PSI is calculated with the flow diagram in Figure 1. A simpler tool, CURB-65, was developed to predict risk for inpatient mortality with only 5 readily available clinical characteristics: presence of confusion, blood urea nitrogen >20 mg/dL, respiratory rate >30, systolic blood pressure <90 or diastolic blood pressure <60, and age >65. An evenly weighted score of <2 is low-risk, a score of 2 is medium-risk, and a score of >2 is high-risk, with 30-day mortality predicted to be as low as 0.6% for no points and as high as 47% for the highest risk strata. Each of these decision tools has its strengths and weaknesses, with the PSI having slightly better ability to discriminate between risk groups, but CURB-65 has far greater ease of use.[9]

The original authors of the PSI suggested that patients in risk classes I and II could safely be treated as outpatients, whereas clinician judgment may play a bigger role in class III. Similarly, patients with CURB-65 scores of 1 to 2 can be treated as outpatients. Both the Infectious Disease Society and the American Thoracic Society recommend the use of either scoring system in their consensus guidelines.[10]

Case Resolution

In the case example, the doctor determines the risk on the basis of both the CURB-65 and the PSI. The patient is calculated to belong to PSI risk class II and to have a CURB-65 score of 0. The doctor decides that the patient can be safely discharged earlier than initially expected. The patient is reassured and sent home with a 7-day course of oral antibiotics and with follow-up with her primary care provider.

Evaluating Need for Isolation for Pulmonary Tuberculosis

Appropriate isolation of patients with tuberculosis is often delayed, and patients without tuberculosis are often isolated unnecessarily; this aggravates the already stressed bed capacity and throughput issues straining hospitals nationwide. In a low endemic area, for example, tuberculosis was identified in only 1 of 92 patients that were isolated.[11] A tool that can discriminate these patients more appropriately would be useful. With this in mind, consider the next case.

A 67-year-old man with no past medical history presents with a 5-day history of productive

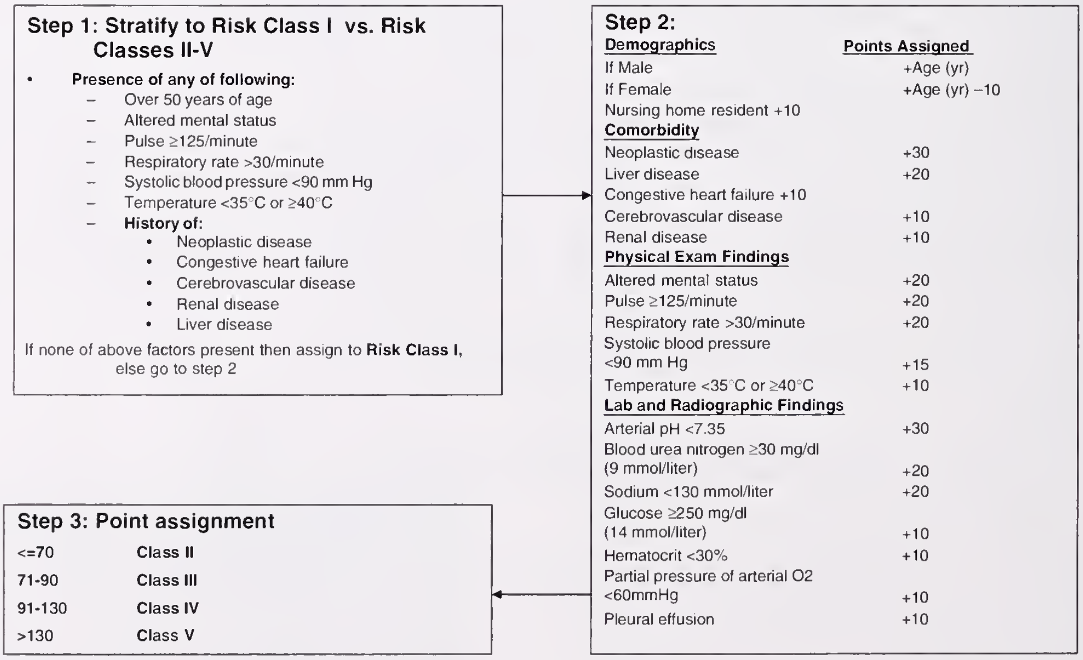

Step 1: Stratify to Risk Class I vs. Risk Classes II-V
Presence of any of following:
Over 50 years of age
Altered mental status
Pulse ≥125/minute
Respiratory rate >30/minute
Systolic blood pressure <90 mm Hg
Temperature <35°C or ≥40°C
History of:
Neoplastic disease
Congestive heart failure
Cerebrovascular disease
Renal disease
Liver disease
If none of above factors present then assign to Risk Class I, else go to step 2
Step 2:
Demographics
Points Assigned
If Male +Age (yr)
If Female +Age (yr) −10
Nursing home resident +10
Comorbidity
Neoplastic disease +30
Liver disease +20
Congestive heart failure +10
Cerebrovascular disease +10
Renal disease +10
Physical Exam Findings
Altered mental status +20
Pulse ≥125/minute +20
Respiratory rate >30/minute +20
Systolic blood pressure <90 mm Hg +15
Temperature <35°C or ≥40°C +10
Lab and Radiographic Findings
Arterial pH <7.35 +30
Blood urea nitrogen ≥30 mg/dl (9 mmol/liter) +20
Sodium <130 mmol/liter +20
Glucose ≥250 mg/dl (14 mmol/liter) +10
Hematocrit <30% +10
Partial pressure of arterial O2 <60mmHg +10
Pleural effusion +10
Step 3: Point assignment
<=70 Class II
71-90 Class III
91-130 Class IV
>130 Class V

Fig 1. Pneumonia Severity Index algorithm. The original authors[8] suggested that patients with class I, class II, and possibly class III pneumonia could be treated as outpatients.

cough, shortness of breath, fever, chills, and scant hemoptysis. Prior to this illness, he was doing well with no report of weight loss or night sweats. Physical examination is unremarkable except for right-sided crackles. An X-ray demonstrates a dense left upper lobe infiltrate. The emergency department physician suspects tuberculosis on the basis of the chest radiograph, and the patient is admitted to the hospitalist service, but no respiratory isolation beds are available.

Several decision tools have been devised to help in the risk stratification of patients with suspected tuberculosis. One, devised by Wisnivesky *et al.*,[12] calculates a score based on 5 readily available clinical criteria, including X-ray findings of upper lobe infiltrate, presence of shortness of breath, fever, a history of positive purified protein derivative, and other tuberculosis risk factors (Table 2). A cutoff of 0 points indicates a need for respiratory isolation. In a prospective study of over 500 patients, the rule demonstrated a negative predictive value of 99.7%. The study estimated that about 35% of potential admissions may have been avoided had the decision rule been used in this cohort of patients.

Other clinical decision rules exist to risk-stratify patients at risk for pulmonary tuberculosis and have been reviewed elsewhere.[11] Self-reported tuberculin skin testing results and upper lobe infiltrates showed the highest association with subsequent diagnosis of infection. Interestingly, hemoptysis, admission white blood cell count, and human immunodeficiency virus status failed to show a predictive association in these studies.

Case Resolution

The presence of an upper lobe infiltrate is of concern (6 points), but the absence of both of a suspicious history for tuberculosis or a positive tuberculin skin test, and the presence of crackles and shortness of breath on examination (−3 points each), gives a total of 0 points, reinforcing a decision not to isolate. This finding with the decision tool reassures the clinician that the patient's symptoms are most consistent with community-acquired pneumonia. The patient is triaged to a regular bed (there are several available), and appropriate treatment for community-acquired pneumonia is provided.

Table 2. *Clinical Prediction Rule for Isolating Patients with Suspected Tuberculosis.*

Variable	Points
Tuberculosis risk factors or symptoms	+4
Positive purified protein derivative tuberculin test results	+5
Shortness of breath	−3
Fever	
<38.5	+0
38.5–39.0	+3
>39.0	+6
Crackles on physical examination	−3
Upper lobe consolidation on chest radiograph	+6

Triage of Patients with Chest Pain

Chest pain accounts for a substantial number of admissions through the Emergency Department. About 5% of patients with acute coronary syndrome (ACS) are improperly diagnosed in the Emergency Department and sent home,[13] and some of these patients go on to experience an adverse cardiac event or arrhythmia. This occurs despite a multitude of ways to risk-stratify patients with clinical or computer algorithms, biomarkers, or continuous ST segment trend monitoring.[14] Consequently, a liberal admission policy for patients with chest pain is common practice and contributes to Emergency Department overcrowding and poor throughput as patients await scarce telemetry beds. Recent studies, however, suggest that a low-risk cohort with suspected ACS can be identified that receives little benefit from inpatient telemetry.[14] The following case illustrates a typical case that may benefit from implementation of a CPR to tease out those patients who would not benefit from a monitored bed.

A 68-year-old man presents to the Emergency Department with severe substernal chest pain radiating to his left arm for 1 hour. He has a history of diabetes but otherwise has an unremarkable past medical history. The pain is relieved with nitroglycerin in the Emergency Department, and the patient is now pain-free. The electrocardiogram is without ischemic changes, and his physical examination is unremarkable. The patient is triaged to a monitored bed with telemetry under the hospitalist service, but there are no monitored beds available.

In 1996, Goldman *et al.*[15] derived a risk index that took into account the following risk factors: presence of ischemic electrocardiographic changes, systolic blood pressure less than 90 mm Hg, congestive heart failure on examination as evidenced by rales above the bases, and presence of pain similar to or worse than prior ACS in a patient with a history of coronary artery disease. The event rates for patients with none of these risk factors were as low as 0.2% in 1 validation study. An impact analysis was performed at Cook County Hospital in Chicago in a prospective cohort of over 1000 patients. Implementing the use of the rule in the triage decisions of the clinicians did not lead to any adverse cardiac events or arrhythmias in unmonitored beds, and more patients with suspected ACS were triaged to an unmonitored bed. In fact, 21% of ACS admissions had been triaged to an unmonitored bed prior to implementation of the rule versus 36% of admissions after its implementation, and this suggests that the use of the Goldman rule may lead to a more efficient use of telemetry resources.[16]

Case Resolution

The patient has no Goldman risk factors present, and this leads the clinician to triage the patient to an unmonitored bed; there are many available. The patient is placed on the schedule for the next available stress test before his discharge is planned.

Validity of Clinical Prediction Rules

CPRs suffer the same limitations as other modern medical interventions: an intervention or rule may work in the narrow confines of a testing environment, but this does not mean it will work in everyday practice.

A hierarchy of evidence (Table 3) provides a framework through which clinicians can judge the strength of evidence in a given CPR.[1] The lowest level (level IV) of evidence is identified when the CPR is derived or created. CPRs are usually derived by the construction of a list of potential predictors, which typically includes items from the history, physical examination, and basic laboratory results. A group of patients are then examined to see if the clinical predictors are present, along with the outcome of interest. Statistical analyses may reveal which predictors are most powerful and which ones can be omitted without the CPR losing its predictive ability.[1] As an example, the PSI was derived in this way by an examination of 250 predictors in over 14,000 patients.[8] Alternatively, clinical predictors may be chosen on the basis of expert consensus much in the way that the Wells criteria were chosen.[6] The rule should make clinical sense and be based on a patient population of sufficient size, with enough outcomes and important predictors present in the study population. Interested readers can find a complete discussion of the derivation of CPRs in an article by Laupacis *et al.*[17]

Even a rigorously derived prediction rule, however, cannot be considered ready for clinical use until the CPR has been prospectively validated so that it can move up the hierarchy to a level II or III rule. The associations found in a level IV rule may be due to chance or may be related to idiosyncratic features of the patient population or the clinicians using the rule.[1] In either case, such a CPR may perform poorly when tested in another patient population or clinical setting. For this reason, validation, or testing the rule in different patient populations with different clinicians and settings, plays a crucial role before it advances up the hierarchy ladder. For example, clinicians will have a higher degree of confidence in the Wells criteria, a level II rule, because they have been validated in over 10 different clinical settings.

Another problem may be that using the rule in everyday clinical scenarios may prove too cumbersome or impractical. For example, the PSI has been criticized because it requires multiple clinical variables before a score can be computed. Some laboratory values may not be available at the time when the emergency clinician is making the decision for inpatient management of the patient's pneumonia. The end result is a CPR that works well in theory but performs poorly in practice. Alternatively, the CPR may be easy to use and accurate but not superior to

Table 3. *Hierarchy of Evidence for Clinical Prediction Rules.*

Level	Rule Category	Underlying Evidence
I	Rules that can be used in a wide variety of settings with confidence that they can change clinician behavior and improve patient outcomes	At least 1 prospective validation in a different population and 1 impact analysis have demonstrated changes in clinician behavior with beneficial consequences.
II	Rules that can be used in various settings with confidence in their accuracy	Accuracy has been either demonstrated in 1 large prospective study involving a broad spectrum of patients and clinicians or validated in several smaller settings that differ from one another
III	Rules that clinicians may consider using with caution and only if the patients in the study are similar to those in their clinical setting	These rules have been validated in only 1 narrow prospective sample.
IV	Rules that need further evaluation before they can be applied clinically	These clinical prediction rules have been derived but not validated or have been validated only in split samples, in large retrospective databases, or by statistical techniques.

clinician intuition, or external forces (eg, fear of litigation or patient demand for services) may override any potential benefit from use of the CPR. For example, fear of litigation for missed myocardial infarction may lead to overdiagnoses of ACS and triage to telemetry despite a reasonably accurate CPR.

For these reasons, an impact analysis study is essential to attain the highest level of evidence for CPRs. An impact study may randomize several institutions to use of the CPR versus several control institutions and measure cost savings, resource utilization, diagnostic accuracy, or safety as outcomes. The complexity of this approach, however, often limits impact analyses to much simpler before-and-after studies at a single institution. The difficulty of carrying out an impact analysis may explain why few CPRs have attained a level I status.

Of the CPRs discussed in the previous examples, only Goldman's CPR has an associated impact analysis that has demonstrated decreased use of telemetry beds with wide implementation of the rule, and so it is the only level I rule. The tuberculosis isolation rule by Wisnivesky *et al.*[12] should be used cautiously as it is a level III rule; it has been validated only in the same institution in which it was derived. The other rules are level II and can be used more confidently.

CONCLUSION

CPRs provide a valuable tool for busy hospitalists to use to incorporate the best evidence into the daily care of their patients. CPRs can help diagnose patients more efficiently, triage them more safely, and help allocate valuable hospital resources safely and rationally. With an understanding of the levels of evidence of CPRs, an informed hospitalist can incorporate other rules into his or her practice as these emerge in the literature. Other CPRs that hospitalists may find useful, along with their associated level of evidence, can be found online at http://www.mssm.edu/medicine/general-medicine/ebm/.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. McGinn TG, Guyatt GH, Wyer PC, et al. Users' guides to the medical literature: XXII: How to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *JAMA* 2000; 284: 79–84.
2. Demonaco NA, Dang Q, Kapoor WN, Ragni MV. Pulmonary embolism incidence is increasing with use of spiral computed tomography. *Am J Med* 2008; 121: 611–617.
3. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet* 2005; 365: 1500–1505.
4. Stein PD, Fowler SE, Goodman LR, et al. Multidetector computed tomography for acute pulmonary embolism. *N Engl J Med* 2006; 354: 2317–2327.
5. Chunilal SD, Eikelboom JW, Attia J, et al. Does this patient have pulmonary embolism? *JAMA* 2003; 290: 2849–2858.
6. Wells PS, Ginsberg JS, Anderson DR, et al. Use of a clinical model for safe management of patients with suspected pulmonary embolism. *Ann Intern Med* 1998; 129: 997–1005.
7. Wells PS, Anderson DR, Rodger M, et al. Excluding pulmonary embolism at the bedside without diagnostic imaging: management of patients with suspected pulmonary embolism presenting to the emergency department by using a simple clinical model and D-dimer. *Ann Intern Med* 2001; 135: 98–107.
8. Fine MJ, Auble TE, Yealy DM, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med* 1997; 336: 243–250.
9. Aujesky D, Auble TE, Yealy DM, et al. Prospective comparison of three validated prediction rules for prognosis in community-acquired pneumonia. *Am J Med* 2005; 118: 384–392.
10. Mandell LA, Wunderink RG, Anzueto A, et al. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin Infect Dis* 2007; 44(suppl 2): S27–S72.
11. Wisnivesky JP, Serebrisky D, Moore C, et al. Validity of clinical prediction rules for isolating inpatients with suspected tuberculosis. A systematic review. *J Gen Intern Med* 2005; 20: 947–952.
12. Wisnivesky JP, Henschke C, Balentine J, et al. Prospective validation of a prediction model for isolating inpatients with suspected pulmonary tuberculosis. *Arch Intern Med* 2005; 165: 453–457.
13. Pope JH, Aufderheide TP, Ruthazer R, et al. Missed diagnoses of acute cardiac ischemia in the emergency department. *N Engl J Med* 2000; 342: 1163–1170.
14. Hollander JE, Sites FD, Pollack CV Jr, Shofer FS. Lack of utility of telemetry monitoring for identification of cardiac death and life-threatening ventricular dysrhythmias in low-risk patients with chest pain. *Ann Emerg Med* 2004; 43: 71–76.
15. Goldman L, Cook EF, Johnson PA, et al. Prediction of the need for intensive care in patients who come to the emergency departments with acute chest pain. *N Engl J Med* 1996; 334: 1498–1504.
16. Reilly BM, Evans AT, Schaider JJ, et al. Impact of a clinical decision rule on hospital triage of patients with suspected acute cardiac ischemia in the emergency department. *JAMA* 2002; 288: 342–350.
17. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA* 1997; 277: 488–494.

SPECIAL FEATURE

Translational Research: Forging a New Cultural Identity

Barry S. Coller, MD

Rockefeller University, New York, NY

More than a decade ago, Dr. Joseph Goldstein called attention to the increasing dissociation between scientific advances and their translation into improved health with his pithy analysis of the biotechnology industry: "1 new gene per day, 1 new company per week, 1 new drug per year."[1] Unfortunately, the gap continues to grow, with increasing concerns about whether the enormous increase in knowledge brought about by the sequencing of the human genome and other scientific advances are being matched by the translational effort. For example, a recent review by the Congressional Budget Office found that the dramatic increase in inflation-adjusted funding of biomedical research since 1970 by the pharmaceutical industry and the National Institutes of Health (NIH), in addition to the influx of capital from the biotechnology industry, has had only a minor impact on the number of truly new drugs approved by the Food and Drug Administration each year.[2] The outlook for the immediate future does not appear to be much brighter, with declining numbers of new drugs being submitted for regulatory approval[3] and the investment community expressing grave concerns about the prospects for both the biotechnology and pharmaceutical industries.[4,5] It is not surprising, therefore, that there has been intense focus on how to successfully bridge the gap between scientific discovery and the development of new strategies to diagnose, treat, and prevent disease; this process is now commonly called *translational research*.

Because the essence of one branch of translational research involves the integration of basic science and clinical medicine, there has been much focus on the importance of physician scientists, whose training provides them with an understanding of both disciplines.[1,6–10] The other branch of translational research, insuring that medical interventions of proven worth are applied correctly and consistently in clinical practice and in the community, is equally important and require the integration of clinical medicine with other disciplines. While this latter integration is also challenging, I will confine my comments to the interface between science and clinical medicine because my own career has been focused in this area. There has also been general recognition of the importance of shifting the translational effort from individuals to teams because multidisciplinary skills are required.[10] The NIH has responded to the challenge by crafting new training and grant awards and developing a loan repayment program; this has culminated in the creation of the Clinical and Translational Science Award (CTSA) program, whose goal is to transform academic institutions and reengineer the clinical research enterprise (http://www.ctsaweb.org). It emphasizes the training of new investigators and the creation of optimal institutional environments for the conduct of translational research. Ultimately, NIH plans to fund 60 CTSAs, creating both local centers of excellence and a consortium to develop national standards. Metrics are being developed to assess the success of the CTSA program over the next several years, but it will require a much longer period of time to truly judge its ultimate impact, and there already is concern about whether it has sufficient resources to achieve its goals. Thus, this is a propitious time for us to analyze the unique features of translational research, searching for insights that can help guide the research and training efforts in CTSAs and more broadly throughout the academic health center community.

Address Correspondence to:

Barry S. Coller, MD
Rockefeller University
New York, NY
Email: collerb@rockefeller.edu

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20064

© 2008 Mount Sinai School of Medicine

DIFFERENCES BETWEEN BASIC SCIENCE AND TRANSLATIONAL RESEARCH

Translational research differs from scientific discovery in both its goals and processes. Thus, scientific discovery starts with the current conceptual model and systematically tests its validity by assessing whether predictions based on the model are borne out by direct experiments. As long as the experiments are designed and conducted properly, new knowledge is obtained; the data either support the current model or demonstrate that it needs to be modified or discarded. Regardless of the result, the goal of adding to the store of human knowledge is achieved. In contrast, translational research springs from the motivation to improve human health; it starts with the articulation of a health need and then involves searching the available body of scientific knowledge to identify discoveries that can be harnessed to address the need. A bidirectional process between bench and bedside then ensues to develop and refine the idea and convert it into a new drug, diagnostic or therapeutic device, or preventive strategy, ultimately resulting in a change in medical practice, commonly requiring regulatory approval. The final goal is not just new knowledge but better health. Because there is always great uncertainty about the impact of any intervention in humans, especially side effects that cannot be predicted from in vitro studies or animal models, and because it is extremely challenging to design experiments to unequivocally assess safety and efficacy, the likelihood of success is unfortunately small. For example, even among drugs that survive the increasingly rigorous preclinical phase of development, only between 10% and 20% go on to achieve regulatory approval.[3] Despite these daunting odds, the quest is still worth the effort because the potential to improve human health may be enormous, and even failed attempts can provide valuable information for the next attempt. Thus, ideally the translational scientist tries to design experiments that have both a basic science hypothesis and a translational hypothesis because if as expected, the translational component fails, the additional basic science knowledge will improve the likelihood that the next translational attempt will be successful.

IMPLICATIONS OF THE NATURE OF TRANSLATIONAL RESEARCH FOR THE DESIGN OF TRAINING PROGRAMS

On the basis of the forgoing analysis, I propose that an individual or a team performing translational research focused on the interface between basic science and clinical medicine requires 3 key skills: (1) the ability to articulate a health need with the precision of a basic science hypothesis, (2) the ability to design a robust and tractable assay (molecular, cellular, or organismal) to identify potential interventions for further development, and (3) the ability to conceptualize from the outset a phase 3 study to assess the safety and efficacy of the proposed new intervention. Thus, training programs in this area of translational research, sometimes called *T-1 translational research*, should be organized to teach these skills through both experiential and didactic methods. There are many parallels to translational research focused on the interface between clinical medicine and the community, which is sometimes called *T-2 translational research*, but for this discipline, the "assay" is likely to be population-based rather than molecular.

Articulating a health need with precision requires both a firm grasp of the state of art of clinical medicine and critical scientific insight into how to construct a hypothesis. By focusing on a specific goal, one can judge its potential impact on improving health versus the effort and resources that will likely be needed to achieve the goal. In my own career, I have built on the knowledge gained from decades of work by others demonstrating a role for platelets in the pathophysiology of acute coronary syndromes and the emerging knowledge of the molecular basis of platelet physiology to construct the hypothesis that blockade of platelet αIIbβ3 receptors inhibiting platelet aggression by agents that would improve the therapy of ischemic cardiovascular disease in patients having, or at risk of having, platelet-mediated thrombosis.[11]

Assay development requires the sophisticated integration of both scientific and medical knowledge because the assay needs to be both robust and biologically meaningful. Assay design is a crucial component of translational research because it allows one to assess the impact of different interventions in search of ones that provide mechanistic insights and pave the way for improved diagnosis, therapy, or prevention. Practical considerations, such as cost, time required to perform, and complexity, are important, especially if the assay is going to be used to screen large numbers of agents (eg, when one is trying to identify monoclonal antibodies or compounds from large libraries). The endpoint must be easy to interpret, and the result with both positive and negative controls must be highly reproducible. Most importantly, despite the simplification required to achieve the aforementioned goals, the assay must be medically and biologically meaningful, by which I mean that if the agent achieves the desired effect in the assay, there is a high likelihood that it will achieve the desired biological or medical effect when administered to an intact animal or human. In my own research, I developed an assay to measure the ability of the platelet $\alpha IIb\beta 3$ receptor to bind fibrinogen by coating small beads with fibrinogen and then macroscopically determining the ability of platelets to agglutinate the beads by binding to the fibrinogen.[12] This permitted us to screen many hundreds of monoclonal antibodies while searching for an antibody that could block the binding of fibrinogen to the $\alpha IIb\beta 3$ receptor on platelets.[13] One of the antibodies identified with this assay (10E5) was used for the prenatal diagnosis of inherited disorders of the $\alpha IIb\beta 3$ receptor[14] (Glanzmann thrombasthenia), and another (7E3)[15] was later modified in collaboration with scientists at Centocor to become the drug abciximab.[11,16] Subsequent modifications of antiplatelet this same assay have been used to both screen libraries of organic molecules to identify new compounds for potential drug development[17] and develop, in collaboration with the scientists at Accumetrics, a series of point-of-care assays to monitor antiplatelet therapy in clinical settings.[18,19]

To successfully complete a translational project, it is necessary to demonstrate the safety and efficacy of the selected intervention in one or more rigorous clinical trials. Thus, the translational scientist requires a knowledge of the steps involved in preclinical development (medicinal chemistry, toxicology, pharmacology, and manufacturing) as well as an understanding of how to design and analyze early phase human studies. In addition, the investigator must have a firm understanding of the bioethical and biostatistical principles that are required to *design a pivotal study to establish safety and efficacy*. Most importantly, the translational scientist needs to understand the regulatory process and, in particular, the criteria for identifying an endpoint for the study that would justify approval for human use. Intimately connected is a requirement for insight into the practical considerations at the interface between academia and the private sector, including how the study endpoint and number of subjects needed to study will affect the length of time required to complete the study and the study's cost. It also requires an appreciation of the goals of the clinical investigators who will recruit the participants and conduct the study and an understanding of both their eagerness to advance the therapy of the diseases that they treat and their concerns about deviating from current, accepted practice. In my own career, I was fortunate to have the opportunity to learn from, and work with, outstanding scientists at Centocor and superb cardiology groups led by Dr. Chip Gold, Dr. Eric Topol, Dr. Rob Califf, Dr. Elliot Antman, and Dr. Eugene Braunwald, who provided me with an in-depth understanding of these vital issues.[11,20–23]

I recognize that many academic leaders may believe that it is unreasonable to expect that an academic translational scientist or team will have all of the knowledge discussed here before embarking on a project. However, I think that the failure to have at least a conceptual understanding of the entire process can result in mistaken priority setting, inappropriate commitment of resources, and investigator frustration and dropout as predictably unrealistic expectations meet the harsh realities. Thus, formal opportunities to learn about how the pharmaceutical, biotechnology, and medical device industries develop drugs and devices should be incorporated into all translational training programs.

Beyond didactic training, however, academic translational scientists require the ability to move effortlessly back and forth between basic science and clinical medicine, and so they need to appreciate how the cultural differences between these disciplines reflect differences in orientation and perspective and how they are reinforced by differences in training, reward structures, and responses to external pressures. I offer a few examples, which are admittedly exaggerated somewhat to emphasize the contrasts, with the goal of helping to sensitize trainees and mentors alike to the differences.

THE CULTURES OF BASIC SCIENCE AND CLINICAL MEDICINE

Need for Immediate Action Versus Avoiding a Rush to Judgment

Physicians are taught to be able to make decisions rapidly, even if all of the data needed are unavailable or ambiguous, whereas basic scientists are taught to reserve judgment until all of the evidence is carefully evaluated. Thus, the exigencies of having to provide medical care to a sick patient in an emergency or when there is insufficient time for a comprehensive evaluation, require that the physician accept uncertainty. In fact, some medical training is specifically designed to provide a preorganized routine to evaluate problems and initiate emergency measures without prolonged thought so as to avoid potentially dangerous delays. Beyond the exigencies of time pressure, however, physicians come to accept that they will not have all of the information that they would like to make a decision. In contrast, basic science training emphasizes the importance of eliminating uncertainty, the dangers of coming to conclusions too rapidly, and the need to wait for compelling evidence from multiple experimental sources and multiple approaches before modifying existing paradigms. The robustness of new discoveries, especially their ability to be replicated by other investigators, is carefully and critically scrutinized prior to acceptance.

Adherence to Standards of Practice Versus Encouragement To Challenge Existing Paradigms

Physicians are taught to adhere to accepted methods of diagnosis and therapy, whereas basic scientists are taught to try to supplant current scientific paradigms. Evidence-based medicine, practice guidelines, and the concept of standard of care all reinforce for the physician the importance of following currently accepted ideas about the proper way to diagnose and treat illness. Moving outside of these norms incurs considerable risks, both from professional colleagues and from patients (and their malpractice attorneys!) should the clinical outcome be unfavorable. In contrast, basic scientists are encouraged to directly challenge current scientific paradigms by bold experiments that provide new mechanistic information that radically transforms the discipline's conceptual model.

Respect for Hierarchy and Expert Authority Versus Encouragement To Critique Accepted Wisdom

Physicians are taught to respect the authority of those with more experience and expertise, whereas basic scientists are taught to be skeptical of authority and to challenge it. Medical training is organized in variably hierarchical structures, depending on the field, and retains many elements of the apprentice system. Although students and junior trainees are encouraged to ask questions about the diagnostic and therapeutic approaches recommended by those above them in the hierarchy, if there is a disagreement, they are expected to carry out the plan approved by the senior members of the team. In contrast, basic scientist training emphasizes that all scientific models undergo continuous modification and refinements, and sometimes they are radically overturned by new evidence. Thus, although it is crucial to study what the current authorities have written and to consider carefully the opinion of one's mentor, it is equally important to keep probing current models for inconsistencies or inadequacies. Thus, the basic scientist's orientation is to continuously look hard for the weak spots in scientific authority.

Errors as Mortal Threats Versus Inevitable Manifestations of the Creative Process

Physicians are taught the importance of developing systems to avoid error because even the slightest mistake may have grave consequences for the patient. Each day they face a minefield of potential errors, such as incomplete differential diagnoses, drug-drug interactions, disease-disease interactions, and diet-drug interactions. Failure to identify any one of these and to react accordingly may result in a patient suffering permanent damage or even death. That is why they are taught to build systems of care that will identify errors at a point where they can still be corrected and thus prevent the error from causing harm to a patient. In contrast, basic scientists learn that no matter how carefully they have analyzed the literature and their own experiments, many of their hypotheses will not be borne out by their experiments. In addition, some of their most novel experiments may lack an important control the first time that they are performed because the need for that particular control may only become apparent after the preliminary data are analyzed. Thus, the iterative nature of scientific discovery and the relative ease of repeating most experiments make error less threatening. Moreover, because so many scientific discoveries derive from paradoxical findings and serendipitous observations, self-confident scientists publicly present their "errors" as the starting point for developing new theories.

Application of Scientific Knowledge Versus Discovery of Scientific Knowledge

Physicians are taught that they must learn medical science so that they can apply it to improve the health of their patients, whereas basic scientists are taught that their job is to create new scientific knowledge. Thus, for the physician, understanding normal physiology and the pathophysiology of disease is required as a means to improving a patient's health, whereas understanding these processes at ever-increasing levels of sophistication is the primary goal of the basic scientist, regardless of the immediate implications for improving health.

Focus on the Unique Versus Focus on the Common

Physicians are trained to look for the unique features in a patient's illness, whereas basic scientists focus on generalizable principles. After their basic education is complete, physicians readily recognize the common manifestations of disease and begin to devote their attention to identifying unusual features of a patient's illness. The atypical manifestations often reflect unusual genetic or environmental influences or the presence of coexisting diseases that modify the expression of the illness. In some cases, the pattern may be so unusual as to suggest an entirely new illness that has never been described. In fact, the physician scientist places high value on studying single patients, knowing that many important discoveries that have advanced medical science have come from careful clinical observations. In contrast, although basic scientists are certainly on the lookout for paradoxical findings, their orientation is to try to discover general principles through replication, without undo focus on single outlier values that may be the result of many factors, including technical errors. Moreover, the emphasis on the need to replicate results to ensure statistical significance breeds skepticism about anecdotal observations.

Uncontrollable Studies Versus Controllable Studies

A physician never has the ability to control all variables when analyzing patients, whereas basic scientists devote enormous effort to devising experiments in which all elements are controlled except the one under study. Thus, variations in genetic background, as well as coexisting diseases, medications, and environmental influences, including nutritional and occupational exposures, all potentially confound the interpretation of studies conducted in humans. The basic scientist's recognition of the vital importance of controlling all variables leads to a fastidious assessment of the entire range of technologies and model systems available to ensure maximal control over variables, even though the selected model may be less directly reflective of, or applicable to, human disease.

Commitment to the Physician's Oath Versus Commitment to the Search for Truth

As an integral part of their contract with society, physicians freely profess an oath to place their patients' interests above their own and to adhere to a high standard of ethical behavior.[24,25] The oath plays a vital role in defining medical professionalism and consciously or unconsciously constrains the physician scientist's conception of the boundaries of acceptable scientific inquiry. Although there have been a number of calls to develop an oath for basic scientists,[26–28] this idea has not met with widespread acceptance, most likely because scientists are concerned about imposing boundaries that may limit the intellectual freedom required for scientific inquiry. Basic scientists commonly point to the self-righting nature of science because truthfulness, the reproducibility of findings among different laboratories, and the need for internal consistency are central elements in the scientific process. These self-righting mechanisms lack an intrinsic moral dimension, however, because the next logical experiment from a scientific standpoint may be one that transcends the boundaries of accepted moral behavior.

Suits and Ties Versus Jeans and T-Shirts

Dress is one of the readily apparent manifestations of a culture and perhaps the clearest differentiator of physicians and basic scientists. Physicians are taught that their dress needs to inspire confidence in their patients or, at a minimum, not be a distraction. Thus, conservative clothes and traditional grooming are expected. Basic scientists have more freedom to choose clothes and grooming methods based on comfort, affordability, and individual expression, and many take full advantage of the opportunity. The physician scientist faces the binary decision each day of selecting clothes reflecting one or the other culture, knowing that basic scientist attire is not welcome on the wards and that physician's attire is likely to produce some raised eyebrows in the laboratory.

Perceptions and Frames of Reference

All human perceptions depend on the frame of reference of the perceiver. For example, to someone standing on earth, the sun appears to revolve around the earth each day, whereas to someone on the sun, the earth appears to make a single rotation on its axis each day. So it is with social and cultural perceptions. Thus, from the frame of reference of the basic scientist, the physician may seem impatient; too accepting of a hierarchical structure, authority figures, and fragmentary evidence; and conservative in thought and dress. From the frame of reference of the physician, the basic scientist may seem overly skeptical and reluctant to come to a conclusion, divorced from medical reality, and inappropriately dismissive of the value of single clinical observations. Layered on these fundamental differences are economic and lifestyle issues that have the potential to exacerbate the irreducible friction between physician scientists and basic scientists. Thus, basic scientists may envy the higher salaries paid to physicians, and physicians may envy the free tuition and stipends that basic scientists receive during their education, commonly leaving them with much less or no educational debt. They also may envy basic scientists' freedom from overnight, weekend, and holiday patient care responsibilities and their greater control of their working hours.

IMPLICATIONS FOR TRAINING AND NURTURING PHYSICIAN SCIENTISTS

Need for Rigorous Scientific Training and Prolonged Mentoring

It is essential that physician scientists conducting translational research at the interface between basic science and clinical medicine have a rigorous basic science experience that is long enough to both learn techniques and develop the requisite intellectual processes needed to construct and test hypotheses. It is especially valuable if the environment is encouraging and supportive enough to help them ride out the inevitable frustrations that accompany the early phases of a research career, frustrations that may have an exaggerated impact on accomplished physicians who are used to clinical success and the satisfaction that comes from the gratitude of appreciative patients. During this time, physician scientists need to be able to devote at least 75% of their time to research, and their clinical activities need to be structured to complement rather than intrude on their training. Practical issues are also crucial, including access to nearby affordable housing and daycare and the availability of child care on weekends and holidays for brief periods of time to permit them to return to the laboratory. Educational debt weighs heavily on trainees' disposable income, their ability to focus on their research, and their perceptions about the economic viability of a career as a physician scientist, and so a loan forgiveness program is vital.

An excellent mentor remains the single most important factor successful physician scientists identify as the essential ingredient in their careers. The introduction of the K series of mentored grants by the NIH was an outstanding initiative that creatively responded to this reality. The K12 program, which is now the centerpiece of the training component in CTSAs, is an excellent addition to the K series, providing focused exposure to the elements of translational research. Unfortunately, however, even at its maximum length of 5 years, it is unlikely to provide sufficient time to reach scientific independence. The most recent statistics from the NIH indicate that MD investigators obtain their first R01 grant, on average, at the age of 43 to 44,[10,29] which is approximately 14 years after most physicians have completed their subspecialty training. Combining a K12 mentored experience of 3 years with either a K08 or K23 mentored grant of 5 years would provide a continuous period of 8 years of mentoring, which is probably the minimum necessary to absorb the knowledge required to compete successfully for an R01. A mentored experience of up to 8 years' duration is especially important if one wishes to attract late bloomers, those physicians who have not had intense exposure to basic science early in their careers but who decide on an investigative career during their clinical training. These individuals bring both their advanced knowledge of the state of art of medical care, and thus the ability to articulate a health need, and an intense commitment that reflects their willingness to make the additional personal and financial sacrifices necessary to develop a career as a physician scientist. Because a physician scientist's research productivity almost always follows a logarithmic pattern early in his or her career, with a variably long nucleation process followed by a period of exponential growth, it is especially important that the mentored training programs be of sufficient length to reach the period of high output. Thus, the failure to provide an extra year or two may seriously jeopardize the trainee's success in reaching independence and continuing in a career in translational research. The lack of return on the extraordinary financial and educational investments made in such individuals is painful to contemplate. Moreover, because they will be expected to compete for R01 grants with PhD and MD-PhD scientists, who commonly have had a total of approximately 8 to 11 years of mentored predoctoral and postdoctoral research experience, it is unrealistic to expect MD physician scientists to compete successfully unless they have mentored training of similar duration. It is important, therefore, to develop mechanisms of support that provide a sense of stability of long-term training funding, from the NIH and/or an academic institution, contingent, of course, on a demonstration of suitable scientific progress.

Need for Realistic Expectations

It is inevitable that some physician scientists in training will grow more comfortable in clinical medicine or basic science over time and choose to focus exclusively on one or the other. When this happens, it is inappropriate to view this outcome as a failure of the training program or to believe that the training has been wasted, because having dual training will enrich the trainee's future contributions, regardless of their emphasis. Thus, expectations about the rate of dropouts from translational research training programs need to be realistic, lest inappropriate pressures and guilt distort the mentor–trainee–program–sponsor relationship.

Need for Novel Programs

Two megatrends increasingly affect translational research, namely, the shift in the structure of the American family from 1 to 2 careers and the need to forge multidisciplinary teams to address complex biological problems with the goal of improving health. Compelling data from the Association of American Medical Colleges demonstrate that the attrition rate among women entering academic medicine careers is much higher than it is among men, most likely reflecting the stress of competing family and career demands (http://www.aamc.org/data/facultyroster/reports.htm). Although not as well documented, the stress of family demands versus career demands probably accounts for a sizable fraction of male attrition as well.

To address both of these trends simultaneously, I propose that in parallel with the programs designed to facilitate the rapid progression of trainees to full independence as laboratory leaders with an R01 grant, academic medical centers develop pathways that will allow junior investigators to work as members of high-quality integrated scientific teams for variably extended periods of time and then encourage them to emerge as fully independent laboratory leaders when their family or other responsibilities allow them to take on all of the administrative and leadership responsibilities of a principal investigator. Such a program would avoid the irrevocable loss of talented young physician scientists from the translational research pool due to their failure to attain independent funding in an arbitrary time frame. This is especially important during the current period when NIH grants are exceedingly competitive. It would also facilitate the creation of stable multidisciplinary teams with enthusiastic and highly motivated junior scientists fresh from outstanding mentored training and eager to make scientific contributions while they continue their growth as physician scientists.

The success of such programs will depend on the ability of academic leaders to recognize and reward the physician scientists working in teams for their contributions, including academic advancement and institutional prestige. Special vigilance will be required to ensure that the junior faculty members are not exploited by senior investigators, are continually encouraged to become independent when the time is ripe, and are aided in the process of achieving independence.

Implications for Medical Education and Basic Science Training

All medical students should be imbued with the spirit of scientific discovery and disabused of the notion that research can be outsourced to PhD scientists in academia and the biotechnology and pharmaceutical industries. Admission processes and curricula should be scrutinized for inadvertent bias against creative students. Medical students should consider themselves clinical scientists, armed at a minimum with precise observation as their scientific instrument, with a professional obligation to both make new discoveries and disseminate that knowledge rapidly to others so that patients all over the world can benefit. Select in-depth examples should be presented to the students of how great basic scientists develop hypotheses, design experiments, develop methods needed to control variables, and struggle to extract the proper interpretation of the results. Parallels should be drawn to the methods by which master clinicians establish diagnoses. If students consider clinical observation an inadequate tool for serious scientific contributions, they should be reminded that Darwin developed the theory of evolution, which overturned all biological conventional wisdom, using only a magnifying glass as his primary scientific instrument and that within the last 35 years, clinicians have played a vital role in the early recognition of infections caused by *Legionella*, human immunodeficiency virus, West Nile virus, and severe acute respiratory syndrome. Finally, to establish the proper framework to encourage scientific discovery by all medical school graduates, the physician's oath should be amended to include a positive obligation to discover so that medicine can progress as rapidly as possible.

Students should be actively encouraged to consider careers as translational scientists at the beginning of medical school by exposure to outstanding translational scientists, group discussions, and career planning advice. The Pasteur Program at Harvard, under the leadership of Dr. Dennis Ausiello, is an outstanding example of such an initiative.[30] Ausiello has also written about the need to seamlessly integrate physics, chemistry, and mathematics into medical biology because translational investigation is increasingly moving to encompass the entire gamut from human populations to atomic scale structure, chemical bonds, and in silico dynamic simulations.[31] Opportunities for extended periods of research should be encouraged, including the excellent year-long mentored programs supported by the NIH Clinical Center, the CTSA programs, the Doris Duke Foundation, the Sarnoff Foundation, and the Howard Hughes Medical Institute. Although beyond the scope of this essay, a review of the goals and structure of MD-PhD programs in light of the evolving needs of translational research is also in order.

Basic scientists who are interested in translational research should be introduced to a variety of important topics, including drug development and approval, protection of human subjects, and clinical trial design. In fact, several CTSA institutions have already developed certificate programs for basic scientists and basic science trainees in translational research. The differences between basic and clinical investigations should be highlighted, including the unique challenges involved in conducting studies involving humans. Ideally, after an appropriate orientation, they should have mentored experiences in clinical environments so that they can gain an understanding of medical decision making and the way in which clinicians evaluate medical evidence.

Implications for Building Translational Research Teams

Building a multidisciplinary team has unique challenges and opportunities. Effective leadership involves sensitizing the members of the team to the differences in perspective, training, and frames of reference of individual members so as to develop a climate of mutual respect and trust. It also requires the leader to ensure effective communication. For example, scientific knowledge and assumptions that are common knowledge to some members of the team but are not known by others need to be explicitly defined, and discipline-specific scientific jargon should be reduced or eliminated. It also requires building consensus on shared values and goals and the most effective paths to achieve the goals. Inviting feedback from all members of the team in group meetings and on an individual basis ensures that frustrations and misperceptions are aired so that they can be addressed directly. Although extra time and effort are required to build such teams, there are intrinsic benefits in having individuals who are new to one's discipline ask "naïve" questions because they often trigger serious contemplation of assumptions that rest on less than solid evidence. Similarly, presenting one's most recent data to individuals who are not knowledgeable about the field requires precise thought and framing the issues with creative metaphors.

CHALLENGES FOR ACADEMIC MEDICAL CENTERS

The CTSA program and the creative new structures and programs developed by academic medical centers to encourage greater communication and collaboration among basic scientists and clinicians offer hope for major improvements. As the discipline of translational research matures, it needs to develop its own unique cultural identity, reinforced by social, interactions at scientific meetings, opportunities to learn from role models, and recognition of exceptional achievement.

One remaining challenge in many academic centers is to create structures that allow physician scientists to practice medicine part time, including provisions for malpractice insurance and appropriate group clinical coverage. A second challenge is to design medical clinical metrics that reward excellent faculty members financially, and with promotion and tenure, that are sensitive to the needs of physician scientists who divide their time between the 2 cultures. For example, creating financial incentives for achieving threshold levels of salary support from clinical reimbursements or from grant support may inadvertently discriminate against those who obtain partial support from each. Similarly, because physician scientists increasingly work as members of teams that cross departmental and discipline boundaries, criteria for promotion and tenure need to reflect this new development, especially if innovative programs like the one that I have proposed for purposefully delaying the time to independence are to succeed.

CALL FOR A NEW SOCIETY OF CLINICAL AND TRANSLATIONAL SCIENCE

To create a new cultural identity, I propose the creation of a new society for clinical and translational science to advance scientific and educational programs in the field. A yearly meeting patterned after the excellent one currently sponsored by the Association for Clinical Research Training (ACRT) and CTSAs that focuses on trainees, their research, and their career development would help to reinforce a sense of community. The meetings should be designed to complement rather than compete with existing subspecialty societies by also presenting the best of best practices in the conduct of clinical and translational science, including new scientific techniques and novel approaches to clinical trial design, biostatistics, bioinformatics, protection of human subjects, research nursing, research pharmacy, regulatory compliance, and technology transfer. Updates from the NIH, the Food and Drug Administration, philanthropic foundations and the pharmaceutical, biotechnology, medical device, and clinical research organization industries would also be valuable, as would a plenary session composed of outstanding examples of clinical and translational science across the entire T1-T2 spectrum. Such a society could also sponsor a journal and provide information to the public and governmental officials on topics important for the continued success of clinical and translational science. The opportunities for scientific exchange and social bonding would reassure physician scientists and others involved in clinical and translational science that they are not alone but rather are members of a community of people who share similar values and goals.

CONCLUSIONS

In his famous 1959 Rede Lecture at Cambridge,[32] C. P. Snow observed that the intellectual life of Western society was being split into 2 polar groups–literary intellectuals and scientists–reflecting 2 cultures separated by a gulf of mutual incomprehension and lack of understanding. He underscored the serious consequences of the failure of these 2 cultures to communicate for developing sound public policy and for addressing the most important international issues of the time, especially the divide between rich and poor nations. He advocated for changes in the educational system to address the crisis, although he recognized that there was unlikely to be a simple or easy solution. The physician scientist stands astride 2 much more closely connected cultures but ones that are distinct enough to require a real effort to bridge. Thus, understanding the unique cultural values of basic science and clinical medicine provides useful insights that can be converted into optimal strategies to train and support translational physician scientists. Bold action is clearly needed, however, because the stakes are very high. The drumbeat of human demographics, the unsustainable rate of growth in the cost of healthcare, and the social justice imperative of alleviating suffering from disease worldwide all cry out for us to bring creativity, good will, passion, and resources to the quest.

ACKNOWLEDGMENTS

This review was supported in part by a Clinical and Translational Science Award (UL1 RR024143) from the National Center for Research Resources (NCRR) of the National Institutes of Health (NIH).

The author thanks the following people for their valuable comments on an earlier version of this article: Dr. Barbara Alving, Dr. Daniel Ausiello, Dr. Mark Babyatsky, Dr. Scott Friedman, Dr. Anthony Hayward, Dr. Kevin Heller, Dr. Kenneth Kaushansky, Dr. James Krueger, Dr. David Nathan, Dr. Mary Klotman, Dr. Paul Klotman, Dr. John Krinsky, Dr. Daniel Rosenblum, Dr. Andrew Schafer, Dr. Sanford Shattil, and Dr. Lawrence Smith.

DISCLOSURES

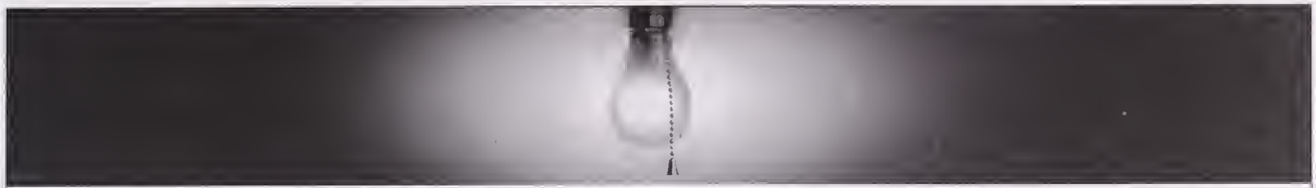
Potential conflict of interest: Barry S. Coller is an inventor of abciximab (Centocor), and in agreement with federal law and the policies of the Research Foundation of the State University of New York, he has a royality interest in abciximab. He is also an inventor of the Verify Now assay system (Accumetrics) and is in accord in federal law and the policies of the Mount Sinai School of Medicine, he has a royalty interest in the VerifyNow assay system (Accumetrics). In addition, he is a consultant to Accumetrics and Novo Nordisk and is an inventor on a pending patent for a new antiplatelet compound.

REFERENCES

1. Goldstein JL, Brown MS. The clinical investigator: bewitched, bothered, and bewildered–but still beloved. *J Clin Invest* 1997; 99: 2803–2812.

DOI:10.1002/MSJ

2. Congressional Budget Office. Research and development in the pharmaceutical industry. Available at: http://www.cbo.gov/ftpdocs/76xx/doc7615/10-02-drugr-d.pdf. 2006; Accessed July 2008.
3. Wood AJ. A proposal for radical changes in the drug-approval process. *N Engl J Med* 2006; 355: 618–623.
4. Pollack, A. It's alive! Meet one of biotech's zombies. *New York Times* February 11, 2007.
5. Martinez B, Goldstein J. Big pharma faces grim prognosis: industry fails to find new drugs to replace wonders like Lipitor. *Wall Street Journal* December 6, 2007.
6. Hauser SL, McArthur JC. Saving the clinician-scientist: report of the ANA long range planning committee. *Ann Neurol* 2006; 60: 278–285.
7. Ley TJ, Rosenberg LE. The physician-scientist career pipeline in 2005: build it, and they will come. *JAMA* 2005; 294: 1343–1351.
8. Nathan DG. Clinical research: perceptions, reality, and proposed solutions. National Institutes of Health Director's Panel on Clinical Research. *JAMA* 1998; 280: 1427–1431.
9. Nathan DG, Schechter AN. NIH support for basic and clinical research: biomedical researcher angst in 2006. *JAMA* 2006; 295: 2656–2658.
10. *Promoting Translational and Clinical Science: The Critical Role of Medical Schools and Teaching Hospitals. Report of the AAMC's Task Force II on Clinical Research, 1–63.* Washington, DC: Association of American Medical Colleges; 2006.
11. Coller BS. Blockade of platelet GPIIb/IIIa receptors as an antithrombotic strategy. *Circulation* 1995; 92: 2373–2380.
12. Coller BS. Interaction of normal, thrombasthenic, and Bernard-Soulier platelets with immobilized fibrinogen: defective platelet-fibrinogen interaction in thrombasthenia. *Blood* 1980; 55: 169–178.
13. Coller BS, Peerschke EI, Scudder LE, Sullivan CA. A murine monoclonal antibody that completely blocks the binding of fibrinogen to platelets produces a thrombasthenic-like state in normal platelets and binds to glycoproteins IIb and/or IIIa. *J Clin Invest* 1983; 72: 325–338.
14. Seligsohn U, Mibashan RS, Rodeck CH, et al. Prenatal diagnosis of Glanzmann's thrombasthenia. *Lancet* 1985; 2: 1419.
15. Coller BS. A new murine monoclonal antibody reports an activation-dependent change in the conformation and/or microenvironment of the platelet GPIIb/IIIa complex. *J Clin Invest* 1985; 76: 101–108.
16. Jordan RE, Wagner CL, Mascelli M, et al. Preclinical development of c7E3 Fab, a mouse/human chimeric monoclonal antibody fragment that inhibits platelet function by blockade of GPIIb/IIIa receptors with observations on the immunogenicity of c7E3 Fab in humans. In: Horton MD, ed. *Adhesion Receptors as Therapeutic Targets.* Boca Raton, FL: CRC Press; 1996; 281–305.
17. Blue R, Murcia M, Karan C, et al. Application of high throughput screening to identify a novel aIIb-specific small molecule inhibitor of £IIbβ3-mediated platelet interaction with fibrinogen. *Blood* 2008; 111: 1248–1256.
18. Coller BS. Monitoring platelet GPIIb/IIIa antagonist therapy. *Circulation* 1998; 97: 5–9.
19. Smith JW, Steinhubl SR, Lincoff AM, et al. Rapid platelet-function assay (RPFA): an automated and quantitative cartridge-based method. *Circulation* 1999; 99: 620–625.
20. Gold HK, Gimple L, Yasuda T, et al. Pharmacodynamic study of F(ab')2 fragments of murine monoclonal antibody 7E3 directed against human platelet glycoprotein IIb/IIIa, in patients with unstable angina pectoris. *J Clin Invest* 1990; 86: 651–659.
21. Coller BS, Anderson K, Weisman HF. Inhibitors of platelet aggregation: GPIIb/IIIa antagonists. In: Braunwald E, ed. *Heart Disease: A Textbook of Cardiovascular Medicine.* Update 4. Philadelphia, PA: W.B. Saunders Co. 1995; 1–10.
22. EPIC Investigators. Use of a monoclonal antibody directed against the platelet glycoprotein IIb/IIIa receptor in high-risk coronary angioplasty. *N Engl J Med* 1994; 330: 956–961.
23. Antman EM, Giugliano RP, Gibson CM, et al. Abciximab facilitates the rate and extent of thrombolysis: results of the thrombolysis in myocardial infarction (TIMI) 14 trial. The TIMI 14 Investigators. *Circulation* 1999; 99: 2720–2732.
24. Miles SM. *The Hippocratic oath and the ethics of medicine.* New York, NY: Oxford University Press; 2005.
25. Coller BS, Klotman P, Smith LG. Professing and living the oath: teaching medicine as a profession. *Am J Med* 2002; 112: 744–748.
26. Mohan-Ram V. 'I swear. . .': an oath for scientists and engineers. *Science* November 10, 2000.
27. Tsai DF, Chen DS. An oath for bioscientists. *J Biomed Sci* 2003; 10: 569–576.
28. Student Pugwash USA. SPUSA pledge. Available at: http://www.spusa.org/pledge. Accessed July 2008.
29. National Institutes of Health. Biomedical research workforce. Available at: http://grants.nih.gov/grants/new_investigators/workforce_info09072007.ppt. Accessed July 2008.
30. What is Pasteur? Available at: http://www.pasteur.harvard.edu. Accessed July 2008.
31. Ausiello D. Science education and communication: AAP presidential address. *J Clin Invest* 2007; 117: 3128–3130.
32. Snow CP. *The Two Cultures.* Cambridge, United Kingdom: Cambridge University Press; 1993.



don't publish online in the dark...

let our open linking network
shed some light.

This journal is CrossRef-ed! As a scientist or scholar navigating digital content, you've come to expect dynamic linking from citation to original source. As an author, you expect ready dissemination of your work to others. CrossRef is a collaborative, cross-publisher reference linking network that allows users to click on a citation and be taken directly to the full-text location. An independent, non-profit organization that employs open standards, CrossRef was established in cooperation with leading research publishers and the International DOI Foundation. Since the service went live in 2000, CrossRef has linked millions of articles from thousands of journals. Find out more at **www.crossref.org**, and make sure all your electronically published work is CrossRef-enabled.

crossref

CROSSREF.ORG | THE CITATION LINKING BACKBONE

200 WHEELER ROAD, 6TH FLOOR, BURLINGTON, MA 01803 • 781-359-2442

▨ **HINARI PHASE I COUNTRIES**
Free access to more than 100 biomedical journals published by Wiley

□ **HINARI PHASE II COUNTRIES**
Nominal cost access to more than 100 biomedical journals published by Wiley

Developing World Access to Research

> "perhaps the biggest step ever taken towards reducing the health information gap between rich and poor countries."

Former WHO Director-General Dr Gro Harlem Brundtland (July, 2001)

Wiley, together with the World Health Organisation and fellow Scientific, Technical and Medical publishers, is proud to play a role in furthering global dissemination of biomedical information.

Through the Health InterNetwork Access to Research Initiative (HINARI) www.healthinternetwork.org, free or nominal cost online access to more than 2,200 biomedical journals from 42 publishers is provided to more than 1,000 institutions in approximately 100 of the world's lowest income countries.

Wiley is one of the six founding publishers who led this initiative with the WHO. We offer more than 100 of our leading biomedical journals through the HINARI program; including many leading journals published with our society partners.

HINARI
Access to Research

WILEY
InterScience®
DISCOVER SOMETHING GREAT
www.interscience.wiley.com

Wiley – helping to make a world of difference.

5756

VOL. 75, NO. 6
NOVEMBER/DECEMBER 2008

MOUNT SINAI JOURNAL OF MEDICINE

A Journal of Translational and Personalized Medicine

Issue Theme

Diversity in Academic Medicine: A Call to Action

Introduction by Gary Butts, MD, Assoc. Dean for Diversity Programs, Policy, and Community Affairs, Mount Sinai School of Medicine

Case for Minority Faculty Development Today

History of Battles Lost and Won

Struggle for Survival among Leading Diversity Programs

Northeast Consortium: Innovation in Minority Faculty Development

Successful Programs in Minority Faculty Development: Overview

Successful Programs in Minority Faculty Development: Ingredients

Special Features

Insulin Strategies for Managing Inpatient and Outpatient Hyperglycemia and Diabetes
Evidence-based Medicine: Case Study in Cessation of Smoking
Prismatic Case: Children with Idiopathic Short Stature

WILEY-BLACKWELL

Discover
Something
Great

Distinctive Content

Content
Journals • Backfile Collections
Reference Works & Encyclopedias
OnlineBooks™
Current Protocols Lab Manuals
Evidence-Based Resources

Access
Single article/chapter or full title
One-time fee, subscription,
or Pay-Per-View options
Unlimited concurrent usage

Capabilities
E-mail Alerts • EarlyView®
Citation download & tracking
Remote access • Mobile Edition®

Promote the process of scientific discovery with a different breed of research system: **Wiley InterScience**®. Long distinguished by its scope and extensive coverage of emerging disciplines, Wiley's authoritative content spans the spectrum of scientific, medical, technical, and business specialties. The dynamic **Wiley InterScience** platform adds value to your research by providing simple desktop access to this preeminent digital library of journals, reference works, books, laboratory manuals, evidence-based databases, and more. Researchers reap the combined benefits of top-tier resources and up-to-the-minute functionality. Experience the difference **Wiley InterScience** can make at your institution today.

WILEY
InterScience®
DISCOVER SOMETHING GREAT

WILEY
1807-2007 KNOWLEDGE FOR GENERATIONS

www.interscience.wiley.com

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

Editor-in-Chief

Penny A. Asbell

Deputy Editors

Dennis S. Charney
Maria New
C. Warren Olanow

Associate Editors

Stuart Aaronson
David H. Adams
Erwin Bottinger
Eric Hollander
Ravi Iyengar
Mary Klotman
Paul Klotman
Ihor R. Lemischka
Fred Lublin
Michael L. Marin
Thomas G. McGinn
Ira Nash
Rosamond Rhodes
Hugh Sampson
Mary Sano
Anthony Squire
Fred Suchy
Savio Woo

Editorial Board

William J. Bremner	University of Washington
Thomas M. Coffman	Duke University
Dwight Evans	University of Pennsylvania
Gabriel Haddad	University of California, San Diego
Heidi E. Hamm	Vanderbilt University
Matti Häyry	University of Manchester
Eve J. Higginbotham	Morehouse School of Medicine
Katherine A. High	University of Pennsylvania
Tony Hope	University of Oxford
John H. Krystal	Yale University
Raju Kucherlapati	Harvard Medical School
Robert J. Lefkowitz	Duke University
Robert Lisak	Wayne State University
George Lister	University of Texas
Philip A. Marsden	University of Toronto
Charles Nemeroff	Emory University
Scott Rauch	Harvard Medical School
Bernard Roizman	University of Chicago
Alan F. Schatzberg	Stanford University
Inder Verma	The Salk Institute
Jeffrey A. Whitsett	University of Cincinnati
Allan Wolkoff	Albert Einstein College of Medicine
Noam Zohar	Bar-Ilan University

Guest Editor

Gary Butts, MD
Associate Dean for Diversity Programs, Policy
and Community Affairs
Mount Sinai School of Medicine

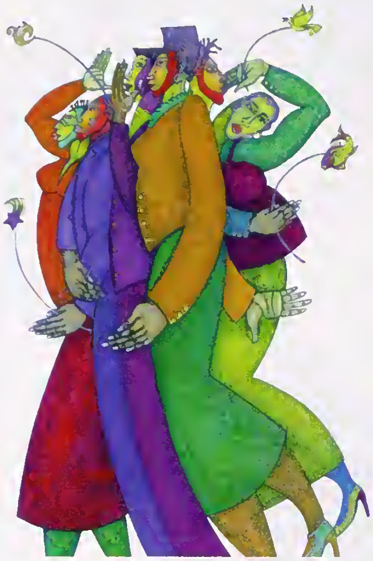
Managing Editor

Stephen Cohen
stephen.cohen@mssm.edu
212-241-6304

Publisher

John Wiley & Sons
Anna Wesley, Associate Editor
awesley@wiley.com

"Diversity" by Charles Bibbs



Charles Bibbs is a contemporary African American artist. Crossing ethnic, gender and generational boundaries, his work conveys a deep sense of spirituality, majesty, dignity, strength, and grace. © 2008 Cover Image by Charles Bibbs.

Themes for future issues of *MSJM*, each shown with Guest Editor

- Advanced Lung Disease (Maria Padilla)
- Traumatic Brain Injury (Andrew Jagoda)
- Zoonotic Infections (Stephanie Factor)
- Medical Education (Lisa Coplit)
- Alzheimer's disease (Nikos Robakis)
- Drug discovery (Ravi Iyengar)
- Personalized Medicine (Erwin Bottinger)
- Neurosurgery (Isabelle Germano)

©2008 **Mount Sinai School of Medicine**. All rights reserved. No part of this publication may be reproduced in any form or by any means, except as permitted under section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the publisher, or authorization through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, Phone: (978) 750-8400; Fax: (978) 750-4770. Such requests and other permission inquiries should be addressed to the Permissions Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; Phone: (201) 748-6011; Fax: (201) 748-6008; E-mail: permreq@wiley.com.

Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine (Print ISSN: 0027-2507; Online ISSN: 1931-7581 at Wiley InterScience, www.interscience.wiley.com) is published twelve times per year for the Mount Sinai School of Medicine by Wiley Subscription Services, a Wiley Company, 111 River Street, Hoboken, NJ 07030.

Send subscription inquiries to: Subscription Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-6645.

Offprint sales and inquiries should be addressed to the Reprint Billing Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-8776. All other inquiries should be addressed to the Customer Service Department at (201) 748-6645.

Individual subscription price: $95.00 in North America, $120.00 outside North America. Payment must be made in US dollars drawn on a US bank. Claims for undelivered copies will be accepted only after the following issue has been delivered. For more subscription prices and information, please visit www.mountsinaijournal.org.

Change of address: Please forward six weeks prior to move: present mailing label with change of address. Please enclose a copy of the mailing label or cite your subscriber reference number in order to expedite handling. Missing copies will be supplied when losses have been sustained in transit and where reserve stock permits. Send claims in care of: *Mount Sinai Journal of Medicine*, c/o John Wiley & Sons, Inc., Attn: Journals Admin Dept. UK, 111 River Street, Hoboken, NJ 07030.

Cancellations: Subscription cancellations will not be accepted after the first issue has been mailed. Periodicals postage paid at Hoboken, NJ and at additional mailing offices.

Postmaster: Send address changes to *Mount Sinai Journal of Medicine*, Subscription Distribution, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Printed in the United States of America.

Advertising inquiries should be addressed to: Advertising Sales Manager, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Phone: (201) 748-6921; Fax: (201) 748-6207. Advertising in this journal is published with a caveat emptor understanding. The Publisher, Editors, and members of the Editorial Board do not imply warranty of products, nor quality, validity, or approval of education material offered by such advertising.

Indexed by: Biological Abstracts® (Thomson ISI), Chemical Abstracts Service/SciFinder (ACS), Current Contents®/Clinical Medicine (Thomson ISI), EMBASE/Excerpta Medica (Elsevier), Index Medicus/MEDLINE/PubMed (NLM), Science Citation Index Expanded™ (Thomson ISI), SCOPUS (Elsevier), Web of Science® (Thomson ISI).

This journal is printed on acid-free paper.

Disclaimer: All articles published, including but not limited to original research, reviews, clinical notes, clinical pearls, prismatic cases, clinician's corner, editorials, reports, and letters, represent the opinions and views of the authors and do not reflect any official policy or medical opinion of the Mount Sinai School of Medicine, the institutions with which the authors are affiliated, or the Publisher unless this is clearly specified. Articles published herein should not be relied upon as recommending or promoting a specific method, diagnosis, or treatment by physicians for any particular patient or disorder. Although the Editors and Publisher believe that drug selections and dosages and the specifications and usage of equipment and devices as set forth herein are in accord with current recommendations and practice at the time of publication, they accept no legal responsibility for any errors or omissions, and make no warranty, express or implied, with respect to material contained herein.

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

CONTENTS
VOL 75, NO 6

Theme Introduction

489 Diversity in Academic Medicine: A Call to Action
Gary Butts

Theme Articles

491 Diversity in Academic Medicine: 1. Case for Minority Faculty Development Today
Marc Nivet, Vera S. Taylor, Gary C. Butts, A. Hal Strelnick, Janice Herbert-Carter, Yvonne W. Fry-Johnson, Quentin T. Smith, George Rust, Kofi Kondwani

499 Diversity in Academic Medicine: 2. History of Battles Lost and Won
A. Hal Strelnick, Elizabeth Lee-Rey, Marc Nivet, Maria L. Soto-Greene

504 Diversity in Academic Medicine: 3. Struggle for Survival Among Leading Diversity Programs
A. Hal Strelnick, Vera S. Taylor, Beverly Williams, Elizabeth Lee-Rey, Janice Herbert-Carter, Yvonne W. Fry-Johnson, Quentin T. Smith, George Rust, Kofi Kondwani

517 Diversity in Academic Medicine: 4. Northeast Consortium: An Innovation in Minority Faculty Development
Gary C. Butts, Jerry Johnson, A. Hal Strelnick, Maria L. Soto-Greene, Beverly Williams, Elizabeth Lee-Rey

523 Diversity in Academic Medicine: 5. Successful Programs in Minority Faculty Development: An Overview
Ann-Gel Palermo, Maria L. Soto-Greene, Vera S. Taylor, Ray Cornbill, Jerry Johnson, Monica Rivera-Mindt, Desiree Byrd, Gary C. Butts, Janice Herbert-Carter, Yvonne W. Fry-Johnson, Quentin T. Smith, George Rust, A. Hal Strelnick

533 Diversity in Academic Medicine: 6. Successful Programs in Minority Faculty Development: Ingredients of Success
Sandra P. Daley, Ann-Gel Palermo, Marc Nivet, Maria L. Soto-Greene, Vera S. Taylor, Gary C. Butts, Jerry Johnson, A. Hal Strelnick, Elizabeth Lee-Rey, Beverly Williams, Jocelyn Dorscher, Wanda D. Lipscomb, Sherria McDowell, Ray Cornbill, Monica Rivera Mindt, Desiree Byrd, Janice Herbert-Carter, Yvonee W. Fry-Johnson, Quentin T. Smith, George Rust, Kofi Kondwani

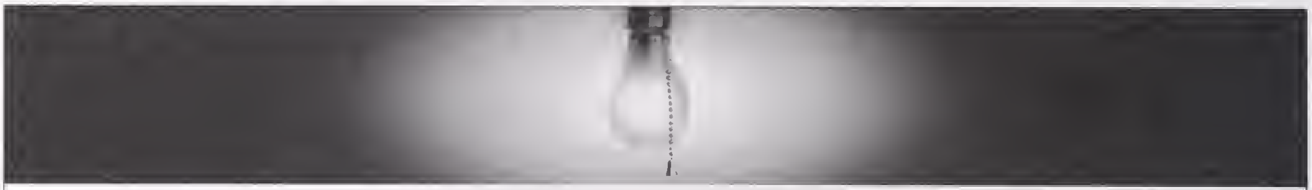
Feature Departments

552 Evidence-based Medicine (Editor: Thomas McGinn)
Case Study in Applying EBM to Current Clinical Problem—Intensive Cessation of Smoking
Deborah Korenstein

556 Prismatic Case (Editor: Maria I. New)
Mutation of Growth Hormone Receptor in Children with Idiopathic Short Stature
Frances M. Guevarra, Maria I. New

Special Feature

558 Insulin Strategies for Managing Inpatient and Outpatient Hyperglycemia and Diabetes
Etie S. Moghissi



don't publish online in the dark...

let our open linking network shed some light.

This journal is CrossRef-ed! As a scientist or scholar navigating digital content, you've come to expect dynamic linking from citation to original source. As an author, you expect ready dissemination of your work to others. CrossRef is a collaborative, cross-publisher reference linking network that allows users to click on a citation and be taken directly to the full-text location. An independent, non-profit organization that employs open standards, CrossRef was established in cooperation with leading research publishers and the International DOI Foundation. Since the service went live in 2000, CrossRef has linked millions of articles from thousands of journals. Find out more at **www.crossref.org**, and make sure all your electronically published work is CrossRef-enabled.

crossref

CROSSREF.ORG | THE CITATION LINKING BACKBONE

200 WHEELER ROAD, 6TH FLOOR, BURLINGTON, MA 01803 • 781-359-2442

THEME INTRODUCTION

Diversity in Academic Medicine: Call to Action

Gary C. Butts, MD

Center for Multicultural and Community Affairs, Mount Sinai School of Medicine, New York, NY

The rationale for increasing diversity in medicine has been supported by increasingly strong evidence. Benefits to the medical educational environment, advances in research for minority populations, improved access to healthcare, and improve patient satisfaction, particularly for diverse populations, have been documented in the published literature.

Nevertheless, despite nearly 50 years of targeted programs and advocacy, the representation of students and faculty from underrepresented minority (URM) groups in US medical schools remains disturbingly inadequate (the Health Resources and Services Administration definition of URM groups includes African and Hispanic Americans, Alaskans, Native Americans, Hawaiians, and certain other Pacific Islanders). Currently, only about 14% of medical students and 7% of faculty in US medical schools are from URM groups.[1,2]

Barriers to increasing URM faculty diversity in US schools of medicine include historical and systemic issues that also serve as deterrents for expanding the pool of URMs (college and medical students) who might declare interest in academic medicine, a critical target for increasing URM faculty. Furthermore, URM faculty appear to be less satisfied than non-URM faculty, and the proverbial glass ceiling appears to be a predominant obstacle for advancement in rank. URM faculty remain noticeably absent in senior ranks and are disproportionately represented at all levels of faculty in medical schools, even when we control for key factors considered important for academic promotion.

Address Correspondence to:

Gary C. Butts, MD
Center for Multicultural and
Community Affairs
Mount Sinai School of Medicine
New York, NY
Email: gary.butts@mssm.edu

The focus of this theme issue of the *Mount Sinai Journal of Medicine*, "Diversity in Academic Medicine: Call to Action," is timely and important. The Association of American Medical Colleges (AAMC), the nation's leading advocate and voice for academic medicine, recently celebrated 30 years of programs in minority affairs. This includes the tireless contributions of the leaders for the Minority Affairs Section of the Minority Affairs Section within the Group on Student Affairs. Recent reorganization under new stewardship at the AAMC and the formation of a new affinity group on diversity present valuable opportunities for the AAMC to reconfirm its mission to diversity and to provide national leadership and guidance for medical schools to increase diversity at all levels of academic medicine, with a particular focus on faculty.

In this theme issue of the *Mount Sinai Journal of Medicine*, the authors present a comprehensive rationale for the importance of URMs in academic medicine; share historical perspectives and challenges to diversity in medicine, including those from minority-serving institutions; describe unique models from successful minority faculty development programs, including the Northeast Consortium for Minority Faculty Development (NECMFD); and detail the ingredients for successful minority faculty development programs.

This theme issue of the *Mount Sinai Journal of Medicine* was conceived by the leaders of the NECMFD, who are nationally recognized for their work in diversity programs and who have aligned their efforts through a partnership funded by the Josiah Macy Jr. Foundation. The NECMFD brings together faculty recruitment and development expertise from 4 urban academic medical centers–the University of Pennsylvania School of

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20085

© 2008 Mount Sinai School of Medicine

Medicine, Mount Sinai School of Medicine, Albert Einstein School of Medicine, and University of Medicine and Dentistry of New Jersey–New Jersey Medical School–all previously funded as Health Resources and Services Administration Centers of Excellence for Minority Health. Through a novel design, this consortium provides an evidence-based approach to advancing URM faculty through a focus on specific skills and areas for professional development.

We thank the editors of the *Mount Sinai Journal of Medicine* for the opportunity to focus on this timely and important issue in academic medicine. We trust that the articles included will prove useful for historical purposes, enlightening readers to struggles and challenges to sustaining momentum for increasing diversity in academic medicine. Additionally, we expect that these articles will serve as a valuable resource for faculty and academic medicine leaders who are interested in designing models or participating in formal faculty development programs. Finally and most importantly, we hope that this theme issue will serve as a rallying call to action for local and national leadership in academic medicine to join forces to support strategies to address the known barriers to increasing diversity in medicine, which have led to clear disparities in the representation of URMs in academic medicine in the United States.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Association of American Medical Colleges. Facts and figures. Available at: http://www.aamc.org/students/minorities/start.htm. Accessed September 2008.
2. Association of American Medical Colleges. Faculty roster 2007. Available at: http://www.aamc.org/data/facultyroster/reports.htm. Accessed September 2008.

Diversity in Academic Medicine No. 1

Case for Minority Faculty Development Today

Marc A. Nivet, EdD,[1] Vera S. Taylor, MSTC,[2] Gary C. Butts, MD,[3] A. Hal Strelnick, MD,[4] Janice Herbert-Carter, MD, MGA, FACP,[5] Yvonne W. Fry-Johnson, MD, MSCR,[6] Quentin T. Smith, MD,[7] George Rust, MD,[6] and Kofi Kondwani, MD[6]

[1]Josiah Macy Jr. Foundation, New York, NY
[2]Morehouse School of Medicine, Atlanta, GA
[3]Center for Multicultural and Community Affairs, Mount Sinai School of Medicine, New York, NY
[4]Institute for Community and Collaborative Health, Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, NY
[5]Department of Medical Education, Morehouse School of Medicine, Atlanta, GA
[6]National Center for Primary Care, Morehouse School of Medicine, Atlanta, GA
[7]Department of Psychiatry and Behavioral Sciences, Morehouse School of Medicine, Atlanta, GA

ABSTRACT

For the past 20 years, the percentage of the American population consisting of nonwhite minorities has been steadily increasing. By 2050, these nonwhite minorities, taken together, are expected to become the majority. Meanwhile, despite almost 50 years of efforts to increase the representation of minorities in the healthcare professions, such representation remains grossly deficient. Among the underrepresented minorities are African and Hispanic Americans; Native Americans, Alaskans, and Pacific Islanders (including Hawaiians); and certain Asians (including Hmong, Vietnamese, and Cambodians). The underrepresentation of underrepresented minorities in the healthcare professions has a profoundly negative effect on public health, including serious racial and ethnic health disparities. These can be reduced only by increased recruitment and development of both underrepresented minority medical students and underrepresented minority medical school administrators and faculty. Underrepresented minority faculty development is deterred by barriers resulting from years of systematic segregation, discrimination, tradition, culture, and elitism in academic medicine. If these barriers can be overcome, the rewards will be great: improvements in public health, an expansion of the contemporary medical research agenda, and improvements in the teaching of both underrepresented minority and non–underrepresented minority students. ***Mt Sinai J Med 75:491–498, 2008.*** © *2008 Mount Sinai School of Medicine*

Key Words: diversity, academic medicine, faculty development program, medical faculty rank, medical faculty tenure, minority faculty, underrepresented minority.

Address Correspondence to:

Marc A. Nivet
Josiah Macy Jr. Foundation
New York, NY
Email: mnivet@macyfoundation.org

Despite almost 50 years of efforts to increase the representation of certain American minorities in the healthcare professions, such representation remains grossly deficient. As defined by the Health Resources and Services Administration, the underrepresented minorities (URMs) include African and Hispanic Americans; Native Americans, Alaskans, and Pacific Islanders (Hawaiians and others); and certain Asians (Hmong, Vietnamese, and Cambodians).[1] In 2006, such minorities composed 28.8% of the American population but only 6% of practicing physicians and 7.2% of full-time medical school faculty.[2,3]

The underrepresentation of URMs in the healthcare professions has a profoundly negative effect on public health.[2,4–11] In 2004, the Sullivan Commission on Diversity in the Healthcare Workforce[11] stated the following:

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20079

© 2008 Mount Sinai School of Medicine

> As the population has become increasingly diverse, glaring disparities in the quality of care, especially for racial and ethnic minorities, have led to thousands of premature deaths each year and incalculable hours of lost productivity, pain, and suffering.

The commission found that, although the health disparities were due in part to a relative lack of health insurance, they were due even more to failure of the healthcare professions to keep pace with changing demographics.

Meanwhile, the percentage of URMs in the American population has continued to increase. The URM percentage of the American population was 27.8% in 2000, 28.8% in 2006, and 31% in 2008.[12] It has been estimated that the URM percentage will be 48% by 2050 (Figure 1).[13,14] Unless URM participation in the healthcare professions increases at the same pace or faster, the URM underrepresentation will get worse.[15]

In its 2004 report,[11] the Sullivan Commission stated that increasing URM representation in the healthcare professions would require an increase in URM medical school administrators and faculty, including both physician and nonphysician scientists.[16–18] Up to that point, the focus of remedial efforts had been on increasing the recruitment and retention of URM medical students. Although agreeing that these efforts were essential, the commission stated that success also required URM faculty and administrators to lead the way, that is, to oversee student and faculty recruitment, retention, and promotion, align an institution's policies with its mission, and set the direction of medical education and curricular reform.[11]

In the remainder of this article, we focus on URM faculty development.

RACIAL AND ETHNIC DISPARITIES IN MEDICAL FACULTY RANK, PROMOTION, AND TENURE

Taking into account both physician and nonphysician scientists, we find that URMs are underrepresented in all ranks of medical school faculty (Table 1). At every level, the faculty are disproportionately white. The disparity in ranking is greatest with respect to full professorships. In 2006, the percentage of full professors who were white was 84%; the percentage of full professors who were URMs was 4.2%.[2,17,19]

URMs and whites are also disparate with respect to tenure (Table 2). For example, in 2006, the percentages of tenured associate and assistant professors who were white, Hispanic, and African American were 80%, 6.7%, and 2.7%, respectively. A similar disparity was found among those on the associate and assistant professor tenure track; of these, the percentages who were white, Hispanic, and African American were 65%, 5.2%, and 3.4%, respectively.[2]

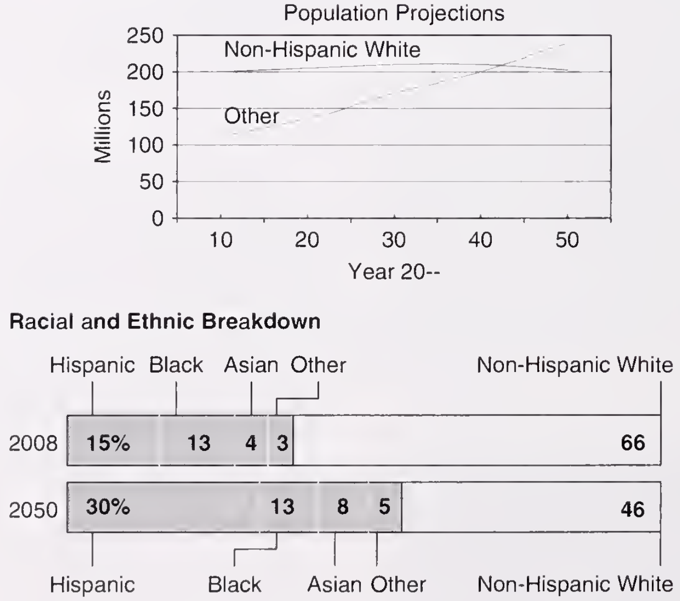


Fig 1. US Census Bureau: Latest projections of changes in the American population with respect to race and ethnicity up to 2050. Adapted from *The New York Times*[14].

Table 1. *Distribution of US Medical School Faculty by Rank and Race/Hispanic Origin.*

Race/Hispanic Origin	Professor	Associate Professor	Assistant Professor	Instructor	Other	Total
Asian	1,999	2,717	8,241	2,377	586	15,920
Black	361	651	2,084	592	90	3,778
Native American/Alaskan	21	16	68	25	6	136
Native Hawaiian/other Pacific Islander	7	16	114	27	10	174
White	24,454	20,192	31,508	8,781	1,850	86,785
Other	25	35	133	33	36	262
Unknown	981	1,253	4,121	1,485	430	8,270
Multiple race	325	411	1,585	435	69	2,825
Cuban	8	10	38	7	3	66
Mexican American	81	128	361	89	27	686
Puerto Rican	173	213	455	91	15	947
Other Hispanic	544	550	1,637	445	59	3,235
Multiple Hispanic	10	35	114	36	1	196
Total	28,989	26,227	50,459	14,423	3,182	123,280

NOTE: Adapted from the 2006 Association of American Medical Colleges faculty roster.[2]

Table 2. *Distribution of US Medical School Faculty by Race/Hispanic Origin, Rank, and Tenure Status.*

Race/Hispanic Origin	Associate Professor (Tenured)	Associate Professor (Tenure Track)	Associate Professor (Not on Track)	Assistant Professor (Tenured)	Assistant Professor (Tenure Track)	Assistant Professor (Not on Track)	Total
Asian	555	417	954	26	1,901	3,597	7,450
Black	138	78	234	9	402	948	1,809
Native American/Alaskan	8	0	8	0	15	22	53
Native Hawaiian/other Pacific Islander	2	2	11	0	27	83	125
White	4,399	2,687	7,666	197	6,653	14,144	35,746
Other	1	5	15	0	34	53	108
Unknown	123	146	369	18	731	1,480	2,867
Multiple race	65	67	178	5	367	835	1,517
Cuban	3	3	2	0	10	16	34
Mexican American	20	10	53	2	54	201	340
Puerto Rican	49	32	57	7	131	170	446
Other Hispanic	87	89	223	2	372	816	1,589
Multiple Hispanic	3	5	17	1	33	44	103
Total	5,453	3,541	9,787	267	10,730	22,409	52,187

NOTE: Adapted from the 2006 Association of American Medical Colleges faculty roster.[2]

The causes of such disparities were investigated by Palepu et al.[20] in a 1998 study involving a stratified random sample of 3013 full-time faculty at 24 US medical schools. This study found that the disparities were consistent even when the researchers controlled for factors that typically influence promotions, such as years as a faculty member, measures of academic productivity, and time spent on clinical activities versus research activities. The authors concluded that the lack of URM promotion in academic medicine might be the result of bias, specifically the stereotype that minorities are inferior. The same question was addressed by Fang et al.[19] in a 2000 study involving 50,145 full-time US medical school faculty who had become assistant or associate professors between 1980 and 1989. This study found that URMs received fewer National Institutes of Health (NIH) awards, and this factor appeared to weigh heavily in decisions concerning faculty promotion. The authors hypothesized that URMs might lag behind whites in faculty promotion because they published less frequently. At the same time, they stated a belief that the lower promotion rate among URMs was not due to a lack of desire or commitment.

BARRIERS AND DETERRENTS TO DEVELOPMENT OF UNDERREPRESENTED MINORITY FACULTY

The development of URM faculty is deterred by barriers resulting from years of systematic segregation, discrimination, tradition, culture, and elitism in academic medicine. These barriers deter both the recruitment of new faculty members, who perceive them in advance, and the retention of current faculty members, who experience them as debilitating and discouraging, often prompting an early departure from academic medicine.[11,19–25] The barriers are discussed next.

Ethnic and Racial Bias and Discrimination

Ethnic and racial bias in academic medicine may be a matter of perception ("not necessarily so") and/or objective demonstration ("undeniably in evidence"). In any case, it results in lower career satisfaction, the correction of which requires effort from all, including both individuals and institutions.[19,21,25,26]

In 2004, Peterson et al.[26] used a 177-item self-administered survey to evaluate the experience of 1979 full-time medical school faculty members at 24 randomly selected US medical schools.[26] The authors found that URM faculty members were more likely than others to perceive ethnic or racial bias in their academic careers and thus to report lower career satisfaction scores. As extensive as the perception of bias was, it was probably more extensive than that reported because the study was limited to current faculty members and did not include those who may have left as a result. In 2005, Price et al.[25] used focus groups and one-on-one semistructured interviews to evaluate the experience of 29 multiethnic faculty members on the tenure track at Johns Hopkins School of Medicine. As perceived by the subjects, their gender, foreign birth, and race or ethnicity provoked subtle manifestations of bias in the promotion process, making residency training and faculty careers more challenging than necessary. Palepu et al.[20] reported that, compared with non-URM faculty members, URM faculty members had significantly lower career satisfaction scores despite similar financial compensation.

To some extent, the burden of discrimination has been inadvertently increased by affirmative action.[21,25,27] Thus, affirmative action has reinforced the belief of some non-URM faculty members that URM faculty members do not irrefutably have the qualifications necessary for an academic career otherwise. Attempting to dispel such negative stereotypes, URM faculty members often feel under intense scrutiny and constant pressure, external or self-imposed.

Isolation and Lack of Networking

Although some URM faculty members work in institutions whose personnel and patients are drawn largely from URMs, most work in institutions whose personnel and patients are largely white. In the latter setting, they work continuously in an environment in which they are few among many. Combined with cultural differences and historical factors, these small numbers tend to isolate URM faculty members within their divisions and departments, reducing their chances of forging personal and professional relationships with non-URM colleagues, sharing information that has a bearing on career development, and gaining recognition for their work.[28] As a result, URM faculty members often report feeling isolated and lonely, and this is a major deterrent to career satisfaction.[19,20]

On the basis of anecdotal evidence, loneliness and isolation are not usually concerns among URM faculty members who work in minority-serving institutions (MSIs). In these settings, they are typically surrounded by faculty and staff who fully understand issues of racial and ethnic bias, health justice, and barriers to the advancement of URM faculty. However, they may find that MSIs have a networking problem of their own: fewer senior faculty with long coat tails, that is, the ability to add junior faculty to grant applications on which they are the principal investigators or to peer-reviewed papers on which they are the lead authors.

Insufficient Time and Energy for Activities That Further Promotion

From the perspective of URM faculty, URM faculty members often have less time than non-URM faculty members for activities that further promotion. Instead of teaching, research, and writing, their time and energy are often consumed by clinical work, contributions to community service, and the mentoring of URM students with complicated academic and personal problems (even though the URM faculty members themselves often lack experience as mentees, let alone training as mentors).[20,29]

Compared with majority-white institutions, MSIs may present URM faculty members with even more demands and less support. Although MSIs have the same administrative and regulatory requirements as larger institutions, they typically have smaller faculties and fewer resources; thus, the burden of committee work, teaching, and clinical service falls heavily on each faculty member, senior or junior. The MSIs may also require junior faculty members to generate research grants and scholarly activity with less institutional support in the form of data, biostatisticians, and technical writers.

Financial Need

Given the barriers to promotion and upward mobility in an academic career, many URM faculty members prefer private practice as a means of paying off their educational debts, and this prompts an early shift from the former to the latter.[25,27]

Lack of Mentoring and Understanding of Requirements for Faculty Success

A lack of mentoring is not unique to URM faculty members, but the effect on URM faculty members is disproportionate. Given the historical weakness in URM faculty development, junior URM faculty members often lack mentors, especially in majority-white institutions. When this is the case, junior faculty members must largely fend for themselves, often failing to gain an understanding of the requirements for faculty success and the realities of faculty appointments and promotion. Even when mentors are available, the mentoring strategies and structures are often insufficient to cope with the many other barriers to the progress of junior members.

Compared with majority-white institutions, MSIs have more mentors but often involve junior faculty in career pathways that lead to only limited success. Thus, many junior faculty members rise rapidly through administrative ranks (student clerkship director, clinical medical director, residency director, etc.) only to find 1 or 2 decades later that, while fulfilling their mid-level leadership roles, they had neglected their scholarly work and lack the research experience and publications necessary to compete for higher level positions such as department chair or dean.

Summary

The faculty experience thus presents a junior URM faculty member with numerous barriers that the member often has no adequate means to overcome. These barriers also engender vicious cycles; for example, in majority-white institutions, a lack of senior faculty members limits the number of available mentors, hindering both recruitment and retention of junior faculty members and thus preventing an increase in the number of senior faculty members in the future. Breaking such vicious cycles and providing junior faculty members with adequate means to overcome the barriers continue to be major challenges.

DEFICIENCIES IN UNDERREPRESENTED MINORITY FACULTY PIPELINE

The medical academic pipeline currently produces only a small trickle of new URM faculty members. For the reasons described previously, deficiencies exist in both recruitment and retention. Although the potential supply of qualified candidates has been increasing, as evidenced by the increasing number of URM students in both undergraduate and graduate education,[23] relatively few URM students become interested in medical careers, and still fewer become interested in academic medicine (Table 3).[3] Moreover, those who enter academic medicine often become discouraged and leave prematurely.

Among the barriers to URM faculty development, the ones most often cited as limiting the flow through the pipeline are the following:[21,25]

- Lack of awareness of academic medicine.
- Lack of role models to engender interest.
- Lack of a welcoming environment, including perceived and/or demonstrable bias and discrimination.
- Lack of diversity and mentors among senior faculty.
- Disillusionment with academic medicine as a career pathway.

In recent years, these barriers have been addressed in new programs aimed at increasing the pipeline flow. Although there have been some apparent successes, the evaluations have been uneven, and the findings often have not been subject to replication. (The findings are discussed in detail in 2 other articles in this theme issue: "Successful Programs in Minority Faculty Development: An Overview" and "Successful Programs in Minority Faculty Development: Essential Ingredients.") Further research to extend, amplify, and validate the existing data is paramount.

Table 3. *Applicants, Accepted Applicants, and First-Year Enrollees (Matriculants) to US Medical Schools in 2006.*

	Race and Ethnicity	Applicants	Accepted Applicants	Matriculants
Hispanic/Latino	Mexican American	955	466	448
	Puerto Rican	560	303	294
	Cuban	206	92	88
	Other Hispanic/Latino	935	415	397
	Multiple Hispanic/Latino	115	59	56
Non-Hispanic/Latino	Black	2,906	1,219	1,155
	Asian	7,532	3,425	3,242
	Native American (including Alaska)	146	70	68
	Native Hawaiian/Pacific Islander	111	33	33
	White	22,670	11,190	10,541
	Other	84	24	23
	Unknown	437	289	276
	Multiple race	925	476	452
Non-US	Foreign	1,527	381	297
TOTAL		39,109	18,442	17,370

NOTE: Adapted from the Web site of the Association of American Medical Colleges.[38]

POTENTIAL CONTRIBUTIONS OF UNDERREPRESENTED MINORITY FACULTY

Improving Public Health

As stated by the US Office of Disease Prevention and Health Promotion, a key objective of Healthy People 2010 (http://www.healthypeople.gov/default.htm) is to overcome ethnic and racial disparities in health and healthcare. These disparities stem in part from cultural and linguistic barriers between minority patients and majority practitioners, a tendency of minority patients to seek care from minority practitioners, a lack of access to services in minority locales, and an insufficient number of minority practitioners.[30,31]

Such barriers to care are likely to grow as the US population becomes more diverse, unless healthcare professions become more diverse and culturally competent at the same time. Cultural competence consists of an understanding of how cultural biases, belief systems, and ethnic origins influence the way in which patients experience illness and respond to treatment.[32] Effective treatment of minority patients requires cultural competence as well as scientific expertise.

Overcoming health disparities thus requires an increase in URM faculty.[8,11,32,33] URM faculty can help attract more URM students to the healthcare professions, increase the number of URM practitioners and thus service in minority locales, and improve the cultural competence of all healthcare practitioners, regardless of their own racial, ethnic, and cultural origins.

Expanding Research Agenda

During the past decade, the NIH has sought to expand biomedical research to take better account of the interests of URMs, specifically by increasing the participation of URM subjects in research studies and by promoting and funding research into questions of special interest to URM health and healthcare.[34]

These objectives can be met more effectively by URM faculty and research scientists than their non-URM counterparts.[32,33] URM faculty and researchers have a better understanding of illnesses and health conditions that disproportionately affect URM and underserved populations and are more likely to design and pursue related research studies.

The NIH has accordingly included URM researchers in its Director's Council of Public Representatives, which advises the NIH on questions affecting public health policy, including the direction of biomedical research.[35]

Improvements in Teaching of Underrepresented Minority and Non–Underrepresented Minority Students

An increase in URM faculty will improve the teaching of all medical students by increasing their cultural competence, particularly through the diversification of case studies and structured dialogues (see the previous discussion of improvements in public health).[36,37]

Compared with non-URM faculty, URM faculty can specifically improve the teaching of URM students by providing them with more frequent student-faculty interactions and greater role modeling, mentorship, and academic guidance.[37]

CONCLUSION: REACHING ETHICAL AND MORAL HIGH GROUND

Increasing the diversity of URM faculty in academic medicine has benefits for both the learning environment in medical schools and the delivery of effective care to diverse populations.

It is also consistent with the ethical and moral principles of the medical profession. In a 1998 editorial in the *Journal of the American Medical Association*, Jordan Cohen, then President of the Association of American Medical Colleges, a nonprofit membership organization composed of medical schools and teaching hospitals in the United States and Canada, stated the following:[21]

> As long as our medical school faculties have little more than token representation from many sectors of the richly diverse American culture, and as long as faculty advancement, for whatever reason, is grossly distorted by race and ethnicity, the medical profession cannot truly lay claim to the ethical and moral high ground it professes to occupy.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. American Association of Medical Colleges. The status of new AAMC definition of "underrepresented in medicine" following the Supreme Court's decision in Grutter. Available at: http://www.aamc.org/meded/urm/statusofnewdefinition.pdf. Accessed September 2008.
2. American Association of Medical Colleges. Faculty roster, December 31, 2006. Available at: http://www.aamc.org. Accessed September 2008.
3. American Association of Medical Colleges. 2007 US medical school entering class is largest ever. Enrollment increases, more black and Hispanic males apply. Available at: http://www.aamc.org/newsroom/pressrel/2007/071016.htm. Accessed September 2008.
4. American Academy of Pediatrics. Committee on Pediatric Workforce. Enhancing the racial and ethnic diversity of the pediatric workforce. *Pediatrics* 2000; 105(1): (pt 1) 129–131.
5. Acosta D, Olsen P. Meeting the needs of regional minority groups: the University of Washington's programs to increase the American Indian and Alaskan native physician workforce. *Acad Med* 2006; 81: 863–870.
6. Grumbach K, Coffman J, Rosenoff E, et al. Strategies for improving the diversity of the health professions (2003). Available at: http://futurehealth.ucsf.edu/pdf_files/strategiesforimprovingfinal.pdf. Accessed September 2008.
7. Institute of Medicine. *Unequal Treatment: What Healthcare Providers Need To Know About Racial and Ethnic Disparities in Healthcare*. Washington, DC: National Academy Press; 2002.
8. Institute of Medicine. *The Nation's Compelling Interest: Ensuring Diversity in the Healthcare Workforce*. Washington, DC: National Academy Press; 2004.
9. Satcher D, Fryer GE Jr, McCann J, et al. What if we were equal? A comparison of the black-white mortality gap in 1960 and 2000. *Health Aff (Millwood)* 2005; 24: 459–464.
10. Simpson CF, Durso Sc, Fried LP, et al. The Johns Hopkins Geriatric Summer Scholars Program: a model to increase diversity in geriatric medicine. *J Am Geriatr Soc* 2005; 53: 1607–1612.
11. Sullivan Commission. Missing persons: minorities in the health professions, a report of the Sullivan Commission on Diversity in the Healthcare Workforce. September 2004. Available at: http://www.aacn.nche.edu/media/pdf/sullivanreport.pdf. Accessed September 2008.
12. US Census Bureau. Table 3: Annual estimates of the population by sex, race, and Hispanic or Latino origin for the United States: April 1, 2000 to July 1, 2006. Release date: May 17, 2007. Available at: http://www.census.gov/popest/race.html. Accessed September 2008.
13. US Census Bureau. An older and more diverse nation by midcentury. Press release, August 14, 2008. Available at: http://www.census.gov/press-release/www/releases/archives/population/012496.html. Accessed September 2008.
14. Roberts S. In a generation, minorities may be the US majority. *New York Times*. August 13, 2008.
15. National Center for Health Workforce Analysis. Changing demographics: implications for physicians, nurses, and other health workers. Available at: http://ftp.hrsa.gov/bhpr/nationalcenter/changedemo.pdf. Accessed October 2007.
16. Johnson J, Williams B, Jayadevappa R. Mentoring program for minority faculty at the University of Pennsylvania School of Medicine. *Acad Med* 1999; 74: 376–379.
17. Daley S, Wingard DL, Reznik V. Improving the retention of underrepresented minority faculty in academic medicine. *J Natl Med Assoc* 2006; 98: 1435–1440.
18. Kosoko-Lasaki O, Sonnino RE, Voytko ML. Mentoring for women and underrepresented minority faculty and students: experience at two institutions of higher education. *J Natl Med Assoc* 2006; 98: 1449–1459.
19. Fang D, Moy E, Colburn L. Racial and ethnic disparities in faculty promotion in academic medicine. *JAMA* 2000; 284: 1085–1092.
20. Palepu A, Carr PL, Friedman RH, et al. Minority faculty and academic rank in medicine. *JAMA* 1998; 280: 767–771.
21. Cohen JJ. Time to shatter the glass ceiling for minority faculty. *JAMA* 1998; 280: 821–822.
22. Cregler LL, Clark LT, Jackson EB Jr. Careers in academic medicine and clinical practice for minorities: opportunities and barriers. *J Assoc Acad Minor Phys* 1994; 5: 68–73.
23. Moody J. *Faculty Diversity: Problems and Solutions*. New York, NY: Routledge Falmer; 2004.
24. Palepu A, Carr PL, Friedman RH, et al. Specialty choices, compensation, and career satisfaction of

underrepresented minority faculty in academic medicine. *Acad Med* 2000; 75: 157–160.

25. Price EG, Gozu A, Kern DE, et al. The role of cultural diversity climate in recruitment, promotion, and retention of faculty in academic medicine. *J Gen Intern Med* 2005; 20: 565–571.
26. Peterson NB, Friedman RH, Ash AS, et al. Faculty self-reported experience with racial and ethnic discrimination in academic medicine. *J Gen Intern Med* 2004; 19: 259–265.
27. Soto-Greene ML, Sanchez J, Churrango J, Salas-Lopez D. Latino faculty development in US medical schools: a Hispanic Center of Excellence perspective. *J Hisp Higher Educ* 2005; 4: 366–376.
28. Coleman J. Social capital and the creation of human capital. *Am J Sociol* 1998; 94: (suppl): S95–S120.
29. Bergen SS. Underrepresented minorities in medicine. *JAMA* 2000; 284: 1138–1139.
30. Fiscella K, Franks P, Gold MR, Clancy CM. Inequality in quality: addressing socioeconomic, racial and ethnic disparities in health care. *JAMA* 2000; 283: 2579–2584.
31. Institute of Medicine. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Healthcare*. Washington, DC: National Academy Press; 2003.
32. Cohen J, Gabriel B, Terrell C. The case for diversity in the health care workforce. *Health Aff* 2002; 21: 90–102.
33. King TE Jr, Dickson TA, DuBose TD Jr, et al. The case for diversity in academic internal medicine. *Am J Med* 2004; 116: 284–289.
34. National Institutes of Health. Research supplements for underrepresented minorities. Available at: http://grants.nih.gov/grants/guide/pa-files/pa-99-104.html. Accessed September 2008.
35. National Institutes of Health. Director's Council of Public Representatives. Available at: http://copr.nih.gov. Accessed September 2008.
36. Milem JF. Attitude change in college students: examining the effect of college peer groups and faculty normative groups. *J Higher Educ* 1998; 69: 117–140.
37. Umbach, P. The contribution of faculty of color to undergraduate education. *Res Higher Educ* 2006; 47: 317–345.
38. Association of American Medical Colleges. Applicants, first-time applicants, acceptees, and matriculants to US medical schools by Hispanic or Latino ethnicity, non-Hispanic or Latino race, and sex, 2005–2007. Facts Table 12, 9/25/2007. Available at: http://www.aamc.org/data/facts/2007/0507sumyrs.htm. Accessed September 2008.

Mount Sinai Journal of Medicine 75:499–503, 2008

Diversity in Academic Medicine No. 2

History of Battles Lost and Won

A. Hal Strelnick, MD[1], Elizabeth Lee-Rey, MD, MPH[2], Marc Nivet, EdD[3], and Maria L. Soto-Greene, MD[4]

[1]Institute for Community and Collaborative Health, Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, NY
[2]Albert Einstein Medical College, Yeshiva University, Bronx, NY
[3]Josiah Macy Jr. Foundation, New York, NY
[4]Hispanic Center of Excellence, New Jersey Medical School, Newark, NJ

ABSTRACT

Spurred by its rapidly changing demographics, the United States is striving to reduce and eliminate racial and ethnic health disparities. To do so, it must overcome the legacy of individual, institutional, and structural racism and resolve conflicts in related political and social ideologies. This has moved the struggle over diversity in the health professions outside the laboratories and ivy-covered walls of academic medicine into the halls of Congress and chambers of the US Supreme Court. Although equal employment opportunity and affirmative action programs began as legal remedies for distinct histories of legally sanctioned racial and gender discrimination, they also became effective means for increasing the representation of underrepresented minorities in higher education and the health professions. Beginning in the 1970s and continuing today, legal challenges to measures for realizing equal opportunity and leveling the playing field have reached the US Supreme Court and statewide ballot initiatives. These historical challenges and successes are the subject of this article. Although the history is not exhaustive, it aims to provide an important context for the struggles of advocates to improve the representation of underrepresented minorities in medicine and reduce racial and ethnic health disparities. ***Mt Sinai J Med 75:499–503, 2008.*** *© 2008 Mount Sinai School of Medicine*

Key Words: academic diversity, academic medicine, affirmative action, anti-affirmative, ethnic diversity, health disparities, pro-affirmative, racial diversity, underrepresented minority.

Address Correspondence to:

A. Hal Strelnick
Institute for Community and
Collaborative Health
Montefiore Medical Center
Albert Einstein College of
Medicine
Bronx, NY
Email: hstrelni@montefiore.org

Rooted in generations of individual, institutional, and structural racism, the long history of minorities pursuing medical education in the United States has many turning points, but perhaps none has been more consequential than the release of the Flexner report in 1910.[1] In that year, Abraham Flexner, an education theorist, was charged by the Carnegie Foundation for the Advancement of Teaching with the task of reviewing all 155 medical schools then in existence in the United States and Canada. The resulting report was a critical exposé of the era's medical education.

Among Flexner's many recommendations, one of the most important was an insistence that medical schools be affiliated with and integrated into an established university structure.[1] Although this and several other reforms recommended by Flexner were credited with greatly improving the quality of American medical education and forcing many for-profit, inadequately financed, and/or poorly managed medical schools to close, they also had the effect of reducing the number of schools preparing physicians to serve disadvantaged communities. The report was particularly critical of the 9 historically black medical colleges, eventually forcing the closure of all but 2, Howard University School of Medicine and Meharry Medical College.

For the next 50 years, these 2 historically black medical colleges were the primary options for African Americans seeking a medical education, and this greatly limited their opportunities. As late as 1964,

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20080

© 2008 Mount Sinai School of Medicine

there were still 5 American medical schools barring African Americans.[2] In that year, 97% of all medical students in the United States were white.[3,4]

START OF EQUAL EMPLOYMENT OPPORTUNITY AND AFFIRMATIVE ACTION PROGRAMS

On March 6, 1961, President John F. Kennedy issued Executive Order 10 925, which created the Committee on Equal Employment Opportunity and mandated that projects financed with federal funds "take affirmative action" to ensure that hiring and employment practices were free of racial bias and discrimination. The first federal sex discrimination legislation, the Equal Pay Act of 1963, responded to the civil rights movement and required equal pay for equal work regardless of sex.[2]

On July 2, 1964, President Lyndon B. Johnson signed the Civil Rights Act of 1964 (Public Law 88–352), which prohibits discrimination in voter registration, public accommodations, employment, and education on the basis of race, color, religion, sex, or national origin. Title VI of the Civil Rights Act forbids discrimination against students on the basis of race, color, or national origin, whereas Title VII of the act, as amended by the Equal Employment Opportunity Act of 1972, forbids employment discrimination on the basic of race, color, national origin, religion, or sex by any employer with 15 or more employees, public and private, with or without federal contracts. For enforcement, the act established the Equal Employment Opportunity Commission.

In 1965, President Johnson issued Executive Order 11 246, requiring all government contractors to take affirmative action toward prospective minority employees in all aspects of hiring and employment and document their efforts to ensure equality in hiring; he amended this in 1967 with Executive Order 11 375 to also cover discrimination on the basis of gender.

ASSOCIATION OF AMERICAN MEDICAL COLLEGES ENTERS THE FRAY

In 1969, the Association of American Medical Colleges (AAMC) received a grant from the US Office of Economic Opportunity to establish an Office of Minority Affairs. In 1970, the AAMC recommended that medical schools achieve equality of opportunity by relieving or eliminating barriers and restraints on access to the medical profession. In April 1970, a task force representing the AAMC, American Medical Association, American Hospital Association, and National Medical Association announced an objective of 12% "representative" minority enrollment for American medical schools by 1975 and recommended substantial increases in and coordination of financial aid to minority students, the creation of an "educational opportunity bank" and a network of regional centers to provide health career counseling for minority students, and an expansion of the AAMC's Office of Minority Affairs.[5,6] Seven years later, a federally commissioned study found that only the last recommendation had been achieved, largely through a $1.5 million US Office of Economic Opportunity grant to the AAMC that in turn supported some 50 programs for minority students from 1969 to 1973.

From 1970 to 1975, aggressive affirmative action admission policies increased the enrollment of underrepresented minority (URM) students from 3% to 10% nationwide.[7] Then, affirmative action programs met with increasing legal opposition. The first anti–affirmative action case, *DeFunis v. Odegaard*, reached the Supreme Court in 1974.[8] The case involved Marcos DeFunis, a student who, having been denied admission to the University of Washington School of Law, sued the school for reverse discrimination. As the case was making its way through hearings and appeals, the student was provisionally admitted. By the time that the case reached the Supreme Court, he was about to graduate. The Court accordingly ruled the case moot.

The second anti–affirmative action case, *Bakke v. Regents of the University of California*, reached the Supreme Court in 1978.[8] In a complex 5-4 decision, the Court ordered the admission of Allen Bakke to the University of California at Davis School of Medicine and declared its separate admissions processes for standard applicants and minority and disadvantaged applicants a violation of the Equal Protection Clause of the US Constitution's 14th Amendment. The Court added that, although race was a legitimate factor to be considered in admissions, fixed quotas were not permissible.

Faced with the stagnation in URM enrollment, the AAMC began a program to overcome the hurdles to a further increase.[9] Then, in 1991, it called for the number of first-year URM medical students to increase to 3000 by the year 2000.[10] Called the "3000 by 2000" initiative, this program fell well short of its goal but succeeded in increasing URM enrollment to more than 12% by 1995.[11] Then, in a sign of retreat, the URM applicant pool decreased from 1998 to 1999 by 6.8%.[12]

During the same decade, the AAMC collaborated with the National Institutes of Health and National Science Foundation to reach the goal of 3000 URM students entering the nation's 125 allopathic medical schools. Although this initiative also fell short of its goal, it produced significant increase in science education partnerships with local school districts, articulation agreements with undergraduate colleges, affiliated health science magnet high schools, classroom and laboratory science enrichment programs in high school and college, and postbaccalaureate programs.

In 1998, the AAMC charged its membership with beginning concerted and systematic efforts to establish effective mentoring and faculty development programs for URM faculty. In his charge, President Cohen stated that "racially and ethnically diverse faculty empowered by the equitable presence of minorities within all ranks of the academy, is the only conceivable bridge to a diverse physician workforce and culturally competent care system."[13]

ATTACKS ON AFFIRMATIVE ACTION INCREASE

In 1995, as President Bill Clinton declared his support for affirmative action programs by announcing the administration's policy of "mend it, don't end it," the Regents of the University of California voted to end affirmative action programs at all University of California campuses. Beginning in 1997 for graduate schools and in 1998 for undergraduate colleges, admissions officers at the University were no longer to use race, gender, ethnicity, or national origin as a factor in their decisions. Meanwhile, in the November 1996 election, California voters narrowly passed Proposition 209, which abolished all public-sector affirmative action programs in employment, education, and contracting in their state. The US Supreme Court, by refusing to hear a challenge to Proposition 209 the following year, allowed it go into effect.[8]

Also in 1996, the US Court of Appeals for the Fifth Circuit ruled against the University of Texas School of Law in the case of *Texas v. Hopwood*, suspending its policy of considering race in the admissions process. Declaring the policy a violation of the Equal Protection Clause, the Court of Appeals stated that "educational diversity is not recognized as a compelling state interest" and held the earlier *Bakke* decision invalid. Because the contested admissions program was no longer in use, the US Supreme Court declined to hear an appeal of the ruling. After the ban on affirmative action, the rates of admission for minority applicants at the flagship institutions of Texas fell sharply. Then, in 1997, the Texas legislature passed the Texas Ten Percent Plan, which ensures that the top 10% of students in all Texas high schools have guaranteed admission to the University of Texas and Texas A&M systems. The effect of the Ten Percent Plan has been to further hinder the admission of minority students to Texas state institutions. However, resisting this anti–affirmative action trend, Houston voters rejected an initiative that would "end affirmative action programs" in city contracting and hiring.

In 1998, Washington state voters approved Initiative 200, which is similar to California's anti–affirmative action Proposition 209. Then, in 2000, the Florida legislature banned affirmative action in higher education as a component of Jeb Bush's "One Florida" initiative and passed a Texas-style Talented Twenty Percent Plan, guaranteeing the top 20% of high-school graduates seats in the University of Florida system. In 2001, California followed suit, enacting a plan that guaranteed admission into the University of California system to the top 4% of all high-school seniors and to the top 12.5% after 2 years outside the system.[8]

In 2003, the Supreme Court rendered opposite decisions in 2 cases involving the University of Michigan.[8] In a pro–affirmative action decision in the case of *Grutter v. Bollinger*, the Court (5-4) held that the university's use of race among other factors in its law school admissions program was constitutional because it furthered "a compelling interest in obtaining an educational benefit that flows from student body diversity."[14] Characterizing the school's evaluation of each applicant as "holistic," the Court found that its program was both narrowly tailored and flexible. In an anti–affirmative action decision in the case of *Gratz v. Bollinger*, the Court (6-3) rejected the undergraduate admissions program at the College of Literature, Science, and the Arts, which granted additional points for race and ethnicity. The Court found that this formula did not provide the "individualized consideration" of each applicant deemed necessary by its precedents. In 2006, Michigan voters passed Proposition 2, which banned affirmative action in state contracting, hiring, and employment.[15]

In 2007, the US Supreme Court issued essentially the same ruling (5-4) in 2 cases, *Parents Involved in Community Schools v. Seattle School District* and *Meredith v. Jefferson County Board of Education*.[8] Although holding that voluntary policies of the Seattle and Louisville school districts to promote diversity in public K-12 schools systems were unconstitutional, it upheld the legality of a narrowly tailored use of race in admissions and reiterated that diversity is a compelling interest in higher education, as established in *Grutter v. Bollinger.*[16] In 2008, opponents of affirmative action succeeded in placing initiatives like Proposition 2 on the official ballots in Colorado and Nebraska.[17]

In 2008, the American Medical Association, which had not prohibited discriminatory practices in its membership until 1964,[18] officially apologized "for its past history of racial inequality toward African-American physicians" and committed itself toward "improving its relationship with minority physicians and to increasing the ranks of minority physicians so that the workforce accurately represents the diversity of America's patients."[19–21]

SUMMARY

Spurred by its rapidly changing demographics, the United States is striving to reduce and eliminate racial and ethnic health disparities. To do so, it must overcome a legacy of individual, institutional, and structural racism and resolve conflicts in related political and social ideologies. This has moved the struggle over diversity in the health professions outside the laboratories and ivy-covered walls of academic medicine into the halls of Congress and chambers of the US Supreme Court.

In his outgoing presidential address in November 2005, the president of the AAMC, Jordan J. Cohen, MD, identified academic medicine's "challenge number 1" to be to "increase the racial and ethnic diversity of the medical profession." He concluded, "To reach the long-term goal of a medical profession that looks more like America, we must forge an effective alliance with the public to address the deep-seated societal ills that lie at the root of the problem."[22]

Although equal employment opportunity and affirmative action programs began as legal remedies for distinct histories of legally sanctioned racial and gender discrimination, they also became effective means for increasing the representation of URMs in higher education and the health professions.[23] Beginning in the 1970s and continuing today, legal challenges to measures for realizing equal opportunity and leveling the playing field have reached the US Supreme Court and state-wide ballot initiatives.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Thelin JR. *A History of American Higher Education.* Baltimore, MD: Johns Hopkins University Press; 2004.
2. Brunner B. Timeline of affirmative action milestones. Available at: http://www.infoplease.com/spot/affirmativetimeline1.html. Accessed September 2007.
3. Raup R, Williams EA. Negro students in medical schools in the United States. *J Med Educ* 1964; 39: 444.
4. Shea S, Fullilove MT. Entry of black and other minority students into U.S. medical schools: historical perspective and recent trends. *NEJM* 1985; 212: 933–940.
5. Association of American Medical Colleges. *Report of the Association of American Medical Colleges Task Force to the Inter-Association Committee on Minority Student Opportunities in Medicine.* Washington, DC: Association of American Medical Colleges; 1970.
6. Schildhaus S, Jaggar FM. *An Exploratory Evaluation of U.S. Medical Schools' Efforts to Achieve Equal Representation of Minority Students.* Washington, DC: Department of Health, Education, and Welfare; 1977.; DHEW Publication (HRA) 78–635.
7. Association of American Medical Colleges. *Minorities in Medical Education: Facts and Figures.* Washington, DC: Association of American Medical Colleges; 2005.
8. Wilcher SJ. The history of affirmative action policies. Available at: http://www.inmotionmagazine.com/aahist.html. Accessed September 2008.
9. Association of American Medical Colleges. *Report of the Association of American Medical Colleges Task Force on Minority Student Opportunities in Medicine.* Washington, DC: Association of American Medical Colleges; 1978.
10. Nickens HW, Ready TP, Petersdorf RG. Project 3000 by 2000–racial and ethnic diversity in U.S. medical schools. *N Engl J Med* 1994; 331: 472–476.
11. Butler WT. Project 3000 by 2000: Progress During Tumultous Times. *Academic Medicine* 1999; 74: 308–309.
12. Barzansky B, Jonas HS, Etzel SI. Educational programs in US medical schools, 1999–2000. *JAMA* 2000; 284: 1114–1120.
13. Cohen JJ. Time to Shatter the Glass Ceilin for Minority Faculty. *JAMA* 1998; 280: 821–822.
14. Terrell C, Beaudreau J. 3000 by 2000 and beyond: next steps for promoting diversity in the health professions. *J Dent Educ* 2003; 67: 1048–1052.
15. Shaw G. Michigan passes initiative to ban affirmative action. Available at: http://www.aamc.org/newsroom/reporter/dec06/michigan.html. Accessed September 2008.
16. Howell WLJ. States, Courts affect race-conscious education Policies. AAMC Reporter, August 2007. http://www.aamc.org/newsroom/reporter/aug07/policies.html. Accessed September 2007.
17. Harris S. More battles looming over affirmative action. Available at: http://www.aamc.org/news-

room/reporter/aug08/affaction.htm. Accessed September 2008.

18. Curtis JL. *Blacks, Medical Schools, and Society*. Ann Arbor, MI: University of Michigan Press; 1971.
19. American Medical Association. AMA apologizes for history of racial inequality and works to include and promote minority physicians. Press release, July 10, 2008. Available at: http://www.ama-assn.org/ama/pub/category/18773.html. Accessed September 2008.
20. Baker RB, Washington HA, Olakanmi O, et al. African American physicians and organized medicine, 1846–1968: origins of a racial divide. *JAMA* 2008; 300: 306–314.
21. Davis RM. Achieving racial harmony for the benefit of patients and communities: contrition, reconciliation, and collaboration. *JAMA* 2008; 300: 323–325.
22. Cohen JJ. The work ahead: AAMC president's address 2005. Available at: http://www.aamc.org/newsroom/pressrel/2005/051106.html. Accessed September 2007.
23. Allen WR, Teranishi R, Dinwiddie G, Gonzalez G. Knocking at freedom's door: race, equity and affirmative action in U.S. higher education. *J Public Health Policy* 2002; 23: 440–452.

Diversity in Academic Medicine No. 3

Struggle for Survival Among Leading Diversity Programs

A. Hal Strelnick, MD,[1] Vera S. Taylor, MSTC,[2] Beverly Williams, MS, JD,[3] Elizabeth Lee-Rey, MD, MPH,[4] Janice Herbert-Carter, MD, MGA,[5] Yvonne W. Fry-Johnson, MD, MSCR,[6] Quentin T. Smith, MD,[7] George Rust, MD,[6] and Kofi Kondwani, MD[6]

[1]Institute for Community and Collaborative Health, Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, NY
[2]Morehouse School of Medicine, Atlanta, GA
[3]Center of Excellence for Diversity in Health Education and Research, University of Pennsylvania School of Medicine, Philadelphia, PA
[4]Albert Einstein Medical College, Yeshiva University, Bronx, NY
[5]Department of Medical Education, Morehouse School of Medicine, Atlanta, GA
[6]National Center for Primary Care, Morehouse School of Medicine, Atlanta, GA
[7]Department of Psychiatry and Behavioral Sciences, Morehouse School of Medicine, Atlanta, GA

ABSTRACT

Since efforts to increase the diversity of academic medicine began shortly after the Civil War, the efforts have been characterized by a ceaseless struggle of old and new programs to survive. In the 40 years after the Civil War, the number of minority-serving institutions grew from 2 to 9, and then the number fell again to 2 in response to an adverse evaluation by the Carnegie Foundation for the Advancement of Teaching. For 50 years, the programs grew slowly, picking up speed only after the passage of landmark civil rights legislation in the 1960s. From 1987 through 2005, they expanded rapidly, fueled by such new federal programs as the Centers of Excellence and Health Careers Opportunity Programs. Encompassing majority-white institutions as well as minority-serving institutions, the number of Centers of Excellence grew to 34, and the number of Health Careers Opportunity Programs grew to 74. Then, in 2006, the federal government cut its funding abruptly and drastically, reducing the number of Centers of Excellence and Health Careers Opportunity Programs to 4 each. Several advocacy groups, supported by think tanks, have striven to restore federal funding to previous levels, so far to no avail. Meanwhile, the struggle to increase the representation of underrepresented minorities in the health professions is carried on by the surviving programs, including the remaining Centers of Excellence and Health Careers Opportunity Programs and new programs that, funded by state, local, and private agencies, have arisen from the ashes. ***Mt Sinai J Med 75:504–516, 2008.*** *© 2008 Mount Sinai School of Medicine*

Address Correspondence to:

A. Hal Strelnick
Institute for Community
and Collaborative Health
Montefiore Medical Center
Albert Einstein College
of Medicine
Bronx, NY
Email: hstrelni@montefiore.org

Key Words: academic diversity, academic medicine, center of excellence, faculty development program, Health Careers Opportunity Program, historically black colleges and universities, minority faculty, minority faculty development, minority-serving institution, underrepresented minority.

The history of efforts to diversify academic medicine began shortly after the Civil War with the founding of 2 schools, Howard University School of Medicine and Meharry Medical College, to provide a medical education to African Americans. Subsequently broadened to include all underrepresented minorities (URMs) (ie, African and Hispanic Americans; Native

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20081

© 2008 Mount Sinai School of Medicine

Americans, Alaskans, and Pacific Islanders including Hawaiians; and Asians including Cambodians, Hmong, and Vietnamese), the efforts have been characterized by a ceaseless struggle of both old and new programs to survive and achieve their goals.

MINORITY-SERVING INSTITUTIONS: OLD GUARD SURVIVES AND ACQUIRES NEW RECRUITS

Minority-serving institutions (MSIs) are those committed to diversifying the health professions by providing special encouragement to URMs, particularly by devoting a significant proportion of their enrollments to URM students. The MSIs include historically black colleges and universities (HBCUs), Hispanic-serving institutions (HSIs), Asian/Pacific Island–serving institutions, and tribal-serving institutions.

Since their inception shortly after the Civil War, the MSIs have produced a disproportionate share of our nation's URM healthcare workforce. They now also employ a disproportionate share of URM faculty among US health professions schools.

Designating Minority-Serving Institutions

Under Title III of the Higher Education Act of 1965, its various amendments, and several other laws related to tribal colleges, the US Department of Education Office of Civil Rights defines MSIs on the basis of either legislation or the percentage of minority student enrollment.[1] Until recently, the minimum URM enrollment was 25%.[2] By this standard, MSIs include all the following:

- HBCU health professions schools, including Howard University School of Medicine (Washington, DC), Meharry Medical College Schools of Dentistry, Medicine, and Pharmacy (Nashville, TN), Charles R. Drew University of Medicine and Science (Los Angeles, CA), Morehouse School of Medicine (MSM; Atlanta, GA), Tuskegee University School of Veterinarian Medicine (Tuskegee, AL), and Xavier University School of Pharmacy (New Orleans, LA).
- The schools of medicine at the University of Colorado (Denver, CO), University of Minnesota–Duluth (Duluth, MN), and University of Washington (Seattle, WA), which serves significant percentages of Native Americans.
- The Sophie Davis School of Biomedical Education in New York City, which serves a large proportion of African American and Hispanic students.
- The University of New Mexico School of Medicine and University of Texas Health Science Center at San Antonio (San Antonio, TX), which serve the Latino population.
- The University of Hawaii John A. Burns School of Medicine (Honolulu, HI), which serves a high percentage of Asians, native Hawaiians, and Pacific Islanders.
- Three medical schools in Puerto Rico: Ponce School of Medicine, Universidad Central del Caribe School of Medicine, and University of Puerto Rico School of Medicine.

However, the Centers for Disease Control and Prevention (CDC) subsequently lowered the minimum enrollment for HSIs to 9% (even though the Hispanic/Latino percentage of the population is 14.6%). The occasion for this change was a cooperative funding agreement between the CDC and Hispanic-Serving Health Professions Schools, Inc. (founded by the National Hispanic Medical Association), which targeted the development of Hispanic/Latino faculty (the only CDC funding agreement specifically targeting such faculty). Under the new criterion, the agreement identified 22 medical schools as HSIs.

According to a search of the literature, Internet, and faculty development grants awarded by the Health Resources and Services Administration (HRSA), the only MSI so far availing itself of this funding mechanism has been MSM.

Founding History: From Few to Many

Following the post–Civil War establishment of Howard University School of Medicine and Meharry Medical College, 7 other medical schools opened their doors to African Americans. However, in 1910, all 9 were the subject of a critical report by Abraham Flexner of the Carnegie Foundation for the Advancement of Teaching. Only the Howard University and Meharry medical schools survived the report; the other 7 were forced to close.

Since then, the MSIs have grown again; now there are 17 schools of medicine and other healthcare professions. The current MSIs are shown in Table 1, together with their founding dates.

Fight for Financial Solvency: Origin of Centers of Excellence

Along the way, many MSIs have had to survive chronic financial difficulties, surviving only with federal assistance. In 1981, Congress passed the Omnibus Budget Reconciliation Act of 1981 (Public Law 97-35), which amended Section 788B of

Table 1. *Minority-Serving Institutions: Founding Dates*

Year	Institution
1868	School of Medicine, Howard University, Washington, DC
1876	Schools of Medicine and Dentistry, Meharry Medical College, Nashville, TN
1927	School of Pharmacy, Xavier University, New Orleans, LA
1944	School of Veterinary Medicine, Tuskegee University, Tuskegee, AL
1950	School of Medicine, University of Puerto Rico, San Juan, PR
1961	School of Medicine, University of New Mexico, Albuquerque, NM
1966	Charles R. Drew University of Medicine and Science, Los Angeles, CA
1973	Sophie Davis School of Medicine, New York, NY
1975	Morehouse School of Medicine, Atlanta, GA
1976	School of Medicine, Universidad Central del Caribe, Bayamón, PR
1981	Ponce School of Medicine, Ponce, PR
1991	Medical Hispanic Center of Excellence, University of Texas Health Sciences Center at San Antonio, San Antonio, TX
1991	Native Hawaiian Center of Excellence, John Burns School of Medicine, University of Hawaii, Honolulu, HI
1992	Native American Center of Excellence, School of Medicine, University of Washington, Seattle, WA
1992	American Indian and Alaska Native Program, School of Medicine, University of Colorado, Denver, CO
1993	Center of American Indian and Minority Health, School of Medicine, University of Minnesota-Duluth, Duluth, MN

the Public Health Service Act. The new law authorized 5-year "advanced financial distress" contracts for 4 HBCUs: Meharry Medical College School of Medicine, Meharry Medical College School of Dentistry, Tuskegee University School of Veterinary Medicine, and Xavier University School of Pharmacy.

While providing assistance, the act also required the HBCUs to develop plans for achieving financial solvency within 5 years. However, they were prevented from doing so by the virtual elimination of other types of federal support assisting minority and other disadvantaged schools and students.[3]

Meanwhile, in 1985, the "excess mortality" and poor health status of minority citizens in the United States were documented in a 7-volume report by the Department of Health and Human Services entitled *Report of the Secretary's Task Force on Black and Other Minorities Health*.[4] Spurred by this report and the persistent inability of HBCUs to achieve financial solvency, Congress passed the Excellence in Minority Health Education and Care Act of 1987 (Public Law 100-87). Under this law, the 4 HBCUs covered by the omnibus act of 1981 became Centers of Excellence (COEs) within the HRSA. The fight for the new law was led by David Satcher, MD, PhD, then President of Meharry Medical College and later Surgeon General, and the Association of Minority Health Professions Schools, which represented all of the HBCUs with health professions schools. In the Senate, the legislation was sponsored by Senators J. Bennett Johnston and John Breaux of Louisiana, Senators Albert Gore and Jim Sasser of Tennessee, Senators Howell Helflin and Richard Shelby of Alabama, and Senator Edward Kennedy of Massachusetts. In the US House of Representatives, the bill's sponsors included Representatives Bill Boner of Nashville, William Nichols of Tuskegee, Lindy Boggs of New Orleans, Louis Stokes of Cleveland, and Mickey Leland of Houston with unanimous support from all 3 states' congressional delegations. While the legislation was pending, the deans of all 4 initially affected schools testified before the House Subcommittee on Health and Environment, which was then chaired by Representative Henry Waxman of California.

Morehouse School of Medicine: Leading Developer of African American Faculty

The nation's leading developer of African American clinical faculty is MSM. Established in 1975 as the School of Medicine at Morehouse College, MSM became an independent institution separate from Morehouse College in 1981. In 1985, it became a 4-year school accredited by the Liaison Committee on Medical Education, and in 1993, it assumed full responsibility for teaching all clinical clerkships.

MSM is dedicated to improving the health and well-being of individuals and communities, increasing the diversity of the health professional and scientific workforce, and addressing primary healthcare needs through programs in education, research, and service, with an emphasis on people of color and the underserved urban and rural populations in Georgia and elsewhere in the nation.

The MSM Faculty Development Program began in 1992 with a grant from the HRSA Bureau of Health Professions through the Title VII Family Medicine Faculty Development Grant Program. At the time, the number of full-time African American academic family physicians nationwide was only 52,[5] and difficulties in recruiting and retaining URM faculty were impeding growth of the MSM Department of Family Medicine. Given the grant, MSM expanded the Faculty Development Program to include faculty and community preceptors in all specialties, including non-MD teaching faculty such as nurse practitioners, nutritionists, and health educators. In 1997, the program came under the aegis of MSM's National Center for Primary Care, and in 2002, the Executive Faculty Development Program was established for URMs from around the nation.

HEALTH CAREERS OPPORTUNITY PROGRAMS AND CENTERS OF EXCELLENCE: FLUCTUATING FORTUNES IN TURBULENT TIMES

Before 1963, the federal government provided only limited support to health professions education through research training grants from the National Institutes of Health (NIH). Responding to the nation's shortage of health personnel (minority and otherwise), Congress enacted the Health Professions Educational Assistance Act of 1963 (Public Law 88–129), which was designed to increase the total number of health professionals through grants for expansion of existing schools and incentives for new schools and through student loans and scholarships.

The new law had mixed results. Although health professions student enrollments, graduates, and practitioners increased significantly, minority representation remained extremely low.

Start of Health Careers Opportunity Programs

Responding to the civil rights movement of the 1960s, universities and the federal government recognized that financial aid was necessary but not sufficient and began to seek to reduce the inequities in access to careers in the health professions by removing the cultural, educational, and other discriminatory barriers that historically discouraged minority students from pursuing health careers.[6] Health professions schools re-examined their admissions policies and redefined their goals not only for minority students but also for low-income and other disadvantaged individuals.[7]

Subsequently, the federal government increased its financial support. In 1972, health professions schools began to receive Special Health Career Opportunity Grants (SHCOGs) under Section 774(b) of the Health Manpower Education Initiative Awards (HMEIA) component of the Comprehensive Health Manpower Training Act of 1971 (Public Law 92–157). Under a formula in the law, the percentage of the total appropriations of HMEIA allocated to SHCOGs was 15%. These funds were granted to health professions schools to stimulate and assist minority and disadvantaged students in undertaking and completing their education for careers in the health professions.

In 1976, the SHCOG program and funding formula were extended through the Health Professions Educational Assistance Act of 1976 (Public Law 94–484). This law also set up a new authority, Section 787, which was similar to the SHCOG program but had its own separate authorization and appropriations, which were no longer based on a formula. Section 787 was entitled the Health Careers Opportunity Program (HCOP) and grew from $14.5 million in 1978 to $24.4 million in 1980.[8] The SHCOG program ended in 1977.

The objectives of both the SHCOG program and HCOP were to

- Identify, recruit, and select persons from disadvantaged backgrounds for education and training in a health profession.
- Facilitate the entry of these individuals into a health professions school.
- Provide counseling and other services designed to help the participants complete their professional education successfully.
- Provide preliminary education designed to help the participants successfully complete the regular course of study for a period before their entry into a health professions school or refer them to such programs.
- Publicize existing sources of financial aid available to students at health professions school or those undertaking the necessary training to qualify them for enrollment in these programs.

The targeted health professions were initially dentistry, medicine, optometry, pharmacy, podiatry, and public health. The Omnibus Budget Reconciliation Act of 1981 (Public Law 97–35) expanded HCOP to include allied health professions (previously supported under Section 798), and in 1986, an additional provision allocated 20% of the appropriations to scholarships for students of exceptional financial need at schools of dentistry, medicine, and osteopathic medicine. This Exceptional Financial Need Scholarship Program has also become a separate line in the health professions training appropriations.

Beginning in 2000, funding priority has been given to comprehensive HCOPs that provide and integrate all the educational services described previously and form partnerships between K-12, community college, college, and postbaccalaureate educational programs and community-based organizations. Program participants have included URMs, women, and whites who are socioeconomically disadvantaged. For HCOP, a person with a disadvantaged background is one who comes from one of the following:

- An environment that has inhibited the person from obtaining the knowledge, skills, and abilities required to enroll in and graduate from a health professions school or from an allied health profession program.
- A family with an annual income below a level based on low-income thresholds according to family size.

During HCOP's first decade (1972–1981), a total of $109 million was awarded to 254 institutions to support 300 projects that assisted more than 40,000 individuals. An analysis conducted by the Association of American Medical Colleges (AAMC) in March 1982 for the Division of Disadvantage Assistance of the HRSA, which administered HCOP awards, found that even when it controlled for Medical College Admission Test scores and grade point averages, both minority and nonminority applicants from HCOP-supported undergraduate schools had significantly higher admissions rates to medical school.[6]

Start of Centers of Excellence

As noted previously, COEs were established by the Excellence in Minority Health Education and Care Act of 1987 (Public Law 100–87). The purpose of COEs was to strengthen the nation's capacity to train URM students in the health professions and develop URM healthcare faculty. The first 4 COEs were HBCUs, that is, the Meharry Medical College School of Medicine, Meharry Medical College School Dentistry, Tuskegee University School of Veterinary Medicine, and Xavier University School of Medicine.

In 1990, this legislation was amended under the Disadvantaged Minority Health Improvement Act (Public Law 101–527) to expand COEs to include all the following categories:

- The original 4 HBCUs.
- Hispanic COEs for institutions training a significant number of Hispanic health professionals.
- Native American COEs for identifying and recruiting Native American students into the health professions.
- Other COEs for health professions schools who have minority enrollments consistent with the national average and show interest in facilitating the education of minority health professionals.

The legislation committed the first $12 million in appropriations to the original 4 HBCUs, with priority thereafter given to Hispanic and Native American COEs for funding above that threshold. In 1992, the Health Professions Education Extension Amendments (Public Law 102–408) further clarified the funding preference, establishing a cap of $12 million for the original 4 HBCUs, dividing additional funding between Hispanic and Native American COEs (60%) and other COEs (40%), and extending eligibility to schools of osteopathic medicine and public health and graduate programs in clinical psychology.

Addition of National Centers of Excellence in Women's Health

In 1996, the Office of Women's Health of the Department of Health and Human Services created a new, integrated system of women's health, one involving a "total transformation from the traditionally fragmented approach of academic health centers."[9]

To this end, the Department of Health and Human Services established National COEs in Women's Health.[10] To receive a Women's Health COE grant award and designation, an institution had to integrate the following key components of the COE model:

- State-of-the-art clinical services addressing all of a woman's needs.
- Public and professional education and training.
- Community outreach and linkages.
- Gender-based research.
- Leadership and career advancement for women in the health sciences.

In 1998, the 6 Women's Health COEs so far funded were required to also develop a specific focus on careers of minority women faculty.[11] By 2005, 14 different health and academic disciplines were associated with 20 COEs, with 6 focused especially on women of color.[12]

Reinforcement of Centers of Excellence and Health Care Opportunity Programs

In 1998, the Health Professions Education Partnership Act reauthorized the health professions training programs under Title VII and consolidated 44 individual programs into 7 general authorities to give HRSA more flexibility in administering the grant programs. The bill continued the COEs and HCOPs as distinct authorities within Title VII and combined the Exceptional Financial Need Scholarships and Financial Assistance for Disadvantaged Health Professions Students programs into the Scholarships for Disadvantaged Students program.

To be designated as a COE under this act, a health professions school was required to

- Have a significant number of URM individuals enrolled or accepted for enrollment in the school.
- Be effective in assisting URM students to complete the educational program and receive a degree.
- Be effective in recruiting URM individuals to enroll in and graduate from the school by providing scholarships and other financial assistance and encouraging URM students from all levels of the educational pipeline to pursue health professions careers.
- Make significant recruitment efforts to increase the number of URM individuals serving in faculty or administrative positions at the school.

Furthermore, to receive a COE grant under Section 736, a school had to meet the following legislative mandates:

- Develop a large, competitive applicant pool through links to institutions of higher education, local school districts, and other community-based entities and establish an educational pipeline for health professions careers.
- Establish, strengthen, or expand programs to enhance the academic performance of URM students attending the school.
- Improve the capacity of the school to train, recruit, and retain URM faculty, including the payment of such stipends and fellowships as the secretary determined necessary.
- Carry out activities to improve the information resources, clinical education, curricula, and cultural competence of the graduates of the school related to minority health issues.
- Facilitate faculty and student research on health issues particularly affecting URM groups, including research on issues related to the delivery of health care.
- Carry out a program to teach students how to provide health services to a significant number of URM patients by training the students at community-based health facilities affiliated with schools.

In March 2005, an HRSA-led collaboration published a major curriculum development project entitled *Transforming the Face of the Health Professions Through Cultural and Linguistic Competence Education: The Role of the HRSA Centers of Excellence*. This publication is currently available on the HRSA Web site.[13]

Criticism and Cutbacks of Centers of Excellence and Health Care Opportunity Programs

In 2005, the Office of Management and Budget (OMB) published its review of the health professions training programs under Title VII, using its performance assessment rating tool, which grades federal programs as effective, moderately effective, adequate, or ineffective. After years of effective ratings for Title VII programs, the OMB concluded that these 40 separate authorities were ineffective, giving them scores ranging from a high of 73% for program management to a low of 13% for program results and accountability.[14,15] As a result, the HRSA administrator, Elizabeth Duke, informed COE and HCOP grantees that the Bush administration would no longer support their programs.

In February 2006, the Congressional Government Accountability Office (GAO) issued a report in response to a request from Senator Michael Enzi (R-WY), then Chairman of the Senate Committee on Health, Education, Labor, and Pensions. Entitled *Health Professions Education Programs: Action Still Needed To Measure Impact*, the report reviewed HRSA's evaluation of the Title VII and VIII (nursing) programs against its overall performance goals and found that these goals did not apply to all of the health professions programs and that HRSA's tracking data were problematic. HRSA was also criticized for failing to publish national supply, demand, and distribution projections for the physician and dentist workforces.[16]

With the OMB and GAO reports as the rationale, the President's budget proposals for fiscal years (FYs) 2006–2009 moved toward totally eliminating all federal funding for all Title VII programs (with the exception of Scholarships for Disadvantaged Students). In pursuing this goal, the Bush administration gave no consideration to several contrary reports, including one from the GAO concluding that HCOP awards were made strictly according to peer review scores,[17] 2 reports posted by the HRSA on its Web site in support of diversity programs,[18,19] and prodiversity literature reviews published by *Academic Medicine* and the California Endowment.[20,21]

In February, 2006, the House and Senate finally and narrowly passed the Labor, Health and Human Services, and Education Appropriations FY 2006 bill in a party-line vote with dramatic cuts to COEs and HCOPs. In the November 2006 election, the majorities in both houses of Congress changed from Republican to Democratic, and the lame-duck Congress failed to pass the FY 2007 appropriations bills. The new Democratic Party–controlled 110th Congress passed the FY 2007 appropriations in February 2007 but without changing the funding levels for any Title VII programs except Geriatric Training Centers, which were supported by the outgoing Chair of the House Labor-HHS Appropriations Subcommittee, Ralph Regula (R-OH).

Both the COEs and HCOPs have been funded during the last 2 FYs at levels that permit only minimal participation from eligible institutions. Specifically, in FY 2005, there were 34 COEs and 74 HCOPs in the US receiving federally funding. However, the COE funding level was reduced from $33.7 million in FY 2005 to $12 million in FY 2006 and FY 2007, and the HCOP funding level was reduced from $35.6 million in FY 2005 to $4 million in FY 2006 and FY 2007. The result is that today only 4 COEs and 4 HCOPs (all HBCUs) receive federal funding, and the remaining programs have lost the opportunity to compete for these grants. Although most COEs and HCOPs survive with dramatic staff and program reductions on limited foundation, state, local, and institutional support, some have closed altogether (eg, the Native Hawaiian COE at the University of Hawaii John A. Burns School of Medicine; Figures 1 and 2).[22]

For FY 2008, the House of Representatives restored the funding levels to $28.4 million for each of the programs. However, the Senate once again appropriated level funds only. At press time, the House and Senate had not reconciled these differences.

DIVERSIFICATION BUTTRESSED BY PRIVATE PIPELINE AND RESEARCH FUNDING

Pipeline Programs

Starting in 1996, an increase in the flow of URM students through the academic pipeline became the goal of the Health Professions Partnerships Initiative (HPPI). Administered by the AAMC, HPPI was funded initially by the Robert Wood Johnson Foundation, which was later joined by the W. K. Kellogg Foundation. During its 10 years of operation, HPPI provided 5-year grants in 3 cohorts to a total of 26 academic health centers. The grants were given to partnerships involving K-12 school systems, colleges, community colleges, community-based organizations, and health professions schools, the last seeking to increase their enrollment of URM students by boosting the students' academic performance at lower levels.[23] HPPI evaluations stressed the importance of starting early, working along the pipeline, and integrating partnership cultures.[24] These partnerships subsequently became a model for comprehensive HCOPs.

The AAMC also administers the Minority Medical Education Program, an intensive 6-week summer enrichment program for promising URM college students. Funded by the Robert Wood Johnson Foundation, the program is hosted by medical schools or medical school–college consortia at 12

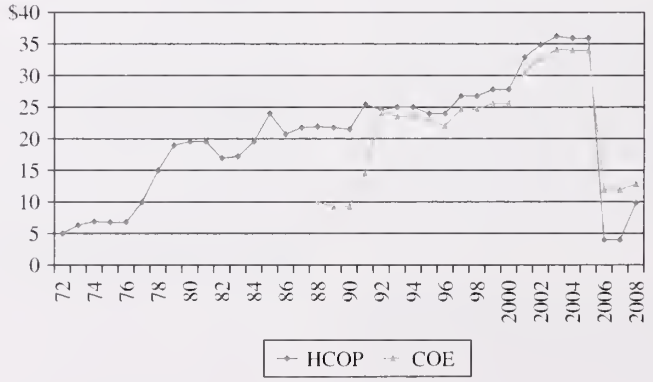


Fig 1. Federal funding history for Centers of Excellence (COEs) and Health Careers Opportunity Programs (HCOPs), 1972–2008 (in millions of dollars unadjusted for inflation). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.].

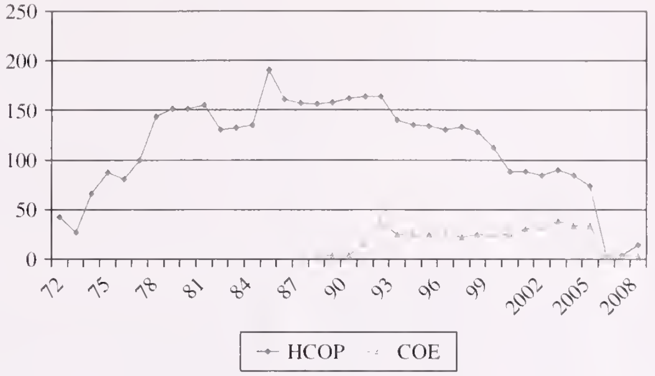


Fig 2. Number of federally funded Centers of Excellence (COEs) and Health Careers Opportunity Programs (HCOPs), 1972–2008. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.].

sites. In 2005, the program evolved into the Minority Medical and Dental Education Program.

Research Programs

In 1987 and 1988, the NIH held a series of US regional meetings to gather testimony on URMs in biomedical and behavioral research. Acknowledging the need to increase the number of URM scientists involved in such research and responding to the concerns expressed at its hearings, the NIH in 1989 created a new administrative mechanism to attract URMs into research at grantee institutions. Targeting URMs from the high-school level to the faculty level, the program eventually came to be called Research Supplements To Promote Diversity in Health-Related Research.[22]

Under the new program, principal investigators of eligible research grants sponsor URM faculty or students at the high-school, undergraduate, graduate, or postdoctoral levels. In FY 2006, some 593 new and 708 continuing awards were made (total = 1301) at the cost of $76.7 million (0.42% of eligible parent award expenditures) with a 75% success rate for applications. New awards peaked in FY 2004 (n = 775), and total awards (n = 1436) and funding ($81.2 million) peaked in FY 2005, whereas the fraction for such supplements as a percentage of total eligible parent award expenditures peaked in FY 1995. In FY 2006, students in high school received 1% of diversity supplement awards (n = 18), and those in college received 6% (n = 83); those in postbaccalaureate, postmasters, predoctoral, and postdoctoral programs received 7% (n = 86), 1% (n = 17), 45% (n = 582), and 26% (n = 330), respectively; and junior faculty investigators received 14% (n = 188). Because the size of an award depends on the career level, increasing from an average of $4900 for high-school students to $111,026 for faculty investigators, expenditures are distributed quite differently, with 36% going to predoctoral students ($26.7 million), 30% to postdoctoral students ($23.3 million), 27% to faculty ($20.9 million), and only 8% ($5.9 million) to the 4 earliest levels in the student pipeline (Figures 3–6).[22]

In 2000, Congress transformed the former Office of Minority Health Research, established in 1990, into a grant-making center within the NIH. Created under the Minority Health and Health Disparities Research and Education Act of 2000 (Public Law 106–525), the new center is called the National Center for Minority Health and Health Disparities (NCMHD). In 2002, the NCMHD funded its first cohort of COEs in Partnerships, Outreach, Research, and Training (Project EXPORT), using 3 different funding mechanisms that reflected different levels of development: Comprehensive Centers (P-60), Exploratory Planning Grants (P-20), and Resource-Related Research Projects (R-24). To secure these grant awards and designations, academic health centers have to demonstrate functional partnerships with minority-serving educational institutions and community-based organizations, research in minority health and health care disparities, research training for minority investigators, and community outreach and information dissemination. In 2007, the Project EXPORT centers evolved into NCMHD Comprehensive Research COEs.

To promote minority health and health disparities research and infrastructure at COEs designated under Section 736 of the Public Health Service Act, the same new law creating the NCHMD established an NCHMD-administered endowment program for COE capacity-building and educational awards. The endowment program thus links 2 separate Department of Health and Human Services agencies, NCMHD and HRSA. To be eligible for NCMHD

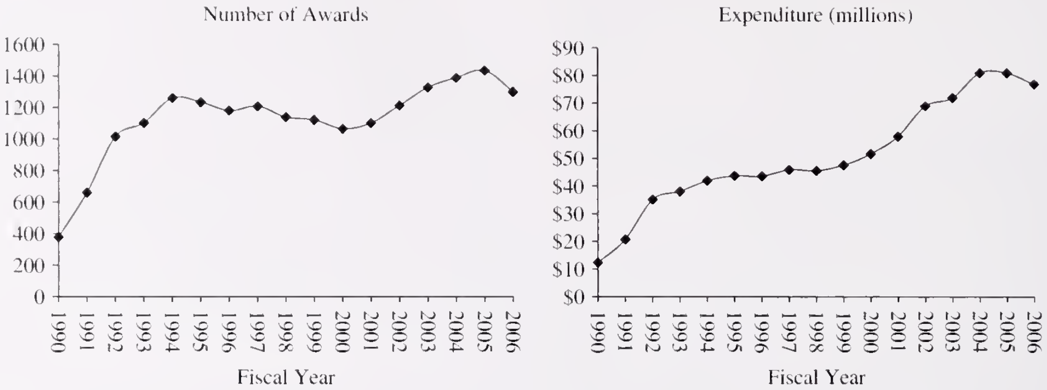
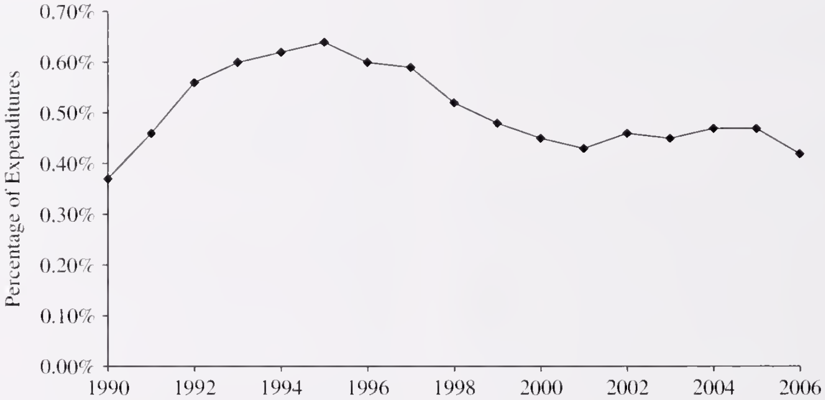


Fig 3. National Institutes of Health supplements to individuals from underrepresented groups or disadvantaged backgrounds: National Institutes of Health–wide trends during fiscal years 1990–2006. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.].



*Eligible grant mechanisms: R01, R03, R10, R15, R18, R21, R22, R24, R35, R37, R41, R42, R43, R44, P01, P20, P30, P40, P41, P50, P51, P60, U01, U10, U19, U41, U42, U54, S06.

Fig 4. National Institutes of Health supplements to individuals from underrepresented groups or disadvantaged backgrounds: National Institutes of Health–wide trends in expenditures as a percentage of expenditures for eligible research grant awards during fiscal years 1990–2006. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.].

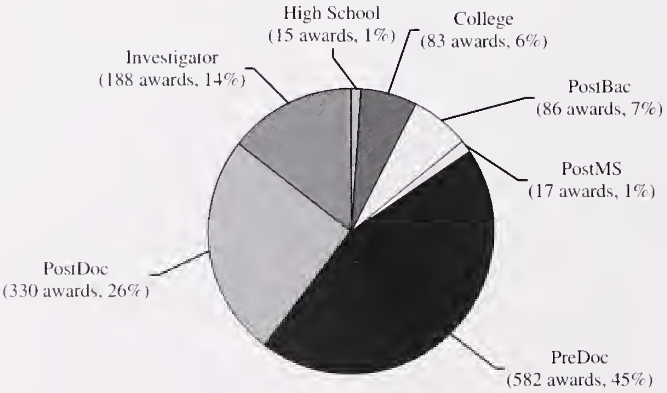


Fig 5. National Institutes of Health supplements to individuals from underrepresented groups or disadvantaged backgrounds: distribution of awards by career level during fiscal year 2006. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.].

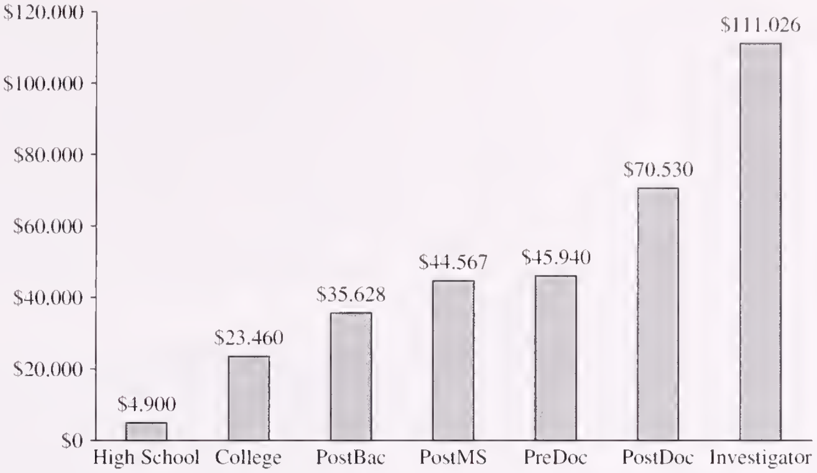


Fig 6. National Institutes of Health supplements to individuals from underrepresented groups or disadvantaged backgrounds: average award size during fiscal year 2006. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.].

endowment funds, a COE must be part of a university whose own endowment falls below the national median.

In FY 2001, the NIH also initiated clinical research loan repayment programs for both individuals from disadvantaged backgrounds and those conducting health disparities research under the administration of NCMHD. Investigators with doctoral degrees and federal, state, or local government research grants working for nonprofit organizations who have student loan debt greater than 20% of their annual salary are eligible for up to $35,000 in loan repayment per year (with coverage of the resulting federal income taxes) for up to two 2-year contracts (or a maximum total of $140,000) if they maintain their research careers. In the first 5 years (FYs 2001–2006), 108 disadvantaged background loans were awarded (55% of applications), and 80 were renewed (71% of applications)[25]; 621 health disparities loans were awarded (51% of applications), and 268 were renewed (66% of applications).[26] Although disadvantaged loan applications and awards have declined, health disparities applications and awards have grown each year since 2001.

In addition to federal programs, the Robert Wood Johnson Foundation offers 4-year postdoctoral research fellowships to URM physicians to support their research and provide them with faculty and National Advisory Committee mentors, thus facilitating their progress and promotion through the ranks as URM faculty. In 2004, the program, which awards up to 12 such fellowships per year, was renamed in honor of Dr. Harold Amos of Harvard University.

ADVOCACY GROUPS IN FRONT LINE, THINK TANKS IN SUPPORT

During the decade from 1995 to 2005, funding for health professions education under Titles VII and VIII of the Public Health Service Act was repeatedly threatened with annual cuts. Because of the chronic and severe nursing shortage, threats to Title VIII funding, which supports nursing education, were not as substantial as those to Title VII, which supports medical education; even so, the threats to both were potentially crippling. For each FY, cuts would be proposed by the president, whether Democratic or Republican, only to be reversed by either the House or Senate Labor-HHS Appropriations Subcommittee after intense lobbying.

After 2005, the nature of the battle changed. In early 2006, Republican fiscal conservatives gained ascendancy in Congress (following the resignation of House Majority Leader Tom Delay [R-TX]) and cut COE and HCOP funding drastically, the former by 81% and the latter by 69%, regardless of any lobbying to the contrary. Then, in late 2006, mid-term elections restored Democrats to majorities in both the House and Senate; this was sufficient to stop any further cuts but, given President Bush's veto, not sufficient to restore COE and HCOP funding to previous levels, so the fight continues.

Health Professions and Nursing Education Coalition

One advocacy group in the forefront of this fight has been the Health Professions and Nursing Education

Coalition (HPNEC), which was organized by the AAMC in 1996.[27] Following the failure of Democratic President Clinton to pass promised healthcare reform legislation in 1994, mid-term elections saw the election of Republican majorities of the House and Senate under the conservative banner, the "Contract with America." Taking account of the new mood, President Clinton proposed annual cuts in Title VII spending with respect to actual spending in the year before; for FY 1995, the cut was 7%, and for FY 1996, the cut was 47%. In December 1995, the new Republican Congress and Democratic president had reached a confrontational impasse on budget appropriations, forcing the federal government to close down all nonessential services until a compromise could be reached. With a Democratic president proposing cuts and a Republican Congress eager to cut domestic programs further still, Title VII funding had seemed in considerable jeopardy. It was then that the AAMC organized national advocates for Title VII and Title VIII funding into HPNEC.

Through advocacy days and legislative breakfasts on Capitol Hill and testimony before and letters to congressional committees, HPNEC members began lobbying key members of Congress and their own representatives for adequate funding. The advocacy effort became even more urgent after the contested election of President Bush in 2000. In comparison with the Clinton administration, the Bush administration proposed even larger annual cuts in Title VII funding, beginning with a 60% cut in its first budget proposal in FY 2001. To lobby against these proposals, HPNEC organized an advocacy day in May 2001 and each year thereafter. In the process, the number of participating organizations from academic medicine and nursing has grown from 15 to 70.

In addition to its advocacy days, HPNEC operates a listserv, circulates letters and petitions, and publishes a detailed brochure citing evidence in support of Title VII and state-by-state funding.

Health Professionals for Diversity Coalition

Another coalition that formed in the mid 1990s was the Health Professionals for Diversity (HPD). Consisting of the health professional schools, national health professions organizations, students, researchers, educators, and policymakers, the HPD focused initially on foiling anti–affirmative action initiatives on state ballots and in the courts; subsequently, it adopted the more general goals of increasing the diversity of health professions and reducing health disparities. Like HPNEC, HPD is staffed by the AAMC.[28]

In 2006, HPD helped to organize Michigan health professional schools to oppose Proposition 2, which banned affirmative action in state contracting, hiring, and employment. Anti–affirmative action advocates had placed the proposition on the Michigan ballot in response to a pro–affirmative action ruling by the Supreme Court. Involving the University of Michigan, the ruling had affirmed that diversity was a compelling interest in higher education.

National Coalition for Diversity in the Health Professions

In 2006, COE directors, previously organized on an informal basis, incorporated themselves as the National Council of Centers of Excellence (NCCOE). Joining forces with the AAMC advocacy group on Capitol Hill, the NCCOE sought to roll back cuts that the Bush administration had made in COE and HCOP funding and otherwise to gain extensions of funds that had not yet been expended (officially called no-cost extensions). Prior to the 2006 mid-term elections, the HRSA offered no-cost extensions of 6 months for HCOPs and 9 months for COEs. Noting that the short term of these extensions would have the effect of ending the grantee status of many COEs and HCOPs before their grants could be formally renewed, the NCCOE sought but failed to gain an extension of 12 months.

After the 2006 mid-term elections, which brought Democratic majorities to both the House and the Senate and renewed hope of restoring COE and HCOP funding to previous levels, the NCCOE reconstituted itself as the National Coalition for Diversity in the Health Professions (NCDHP). Broadening its mission, the NCDHP hired a Washington representative and expanded its alliances. In addition to the AAMC, its allies now included the American Medical Student Association, Association of American Indian Physicians, Association of Minority Health Professions Schools, Hispanic-Serving Health Professions Schools, National Association of Community Health Centers, National Association of Minority Medical Educators, National Center of Minority Health, National Dental Association, National Education Association, National Hispanic Medical Association, and National Medical Association.

Strengthened by these alliances, the NCDHP finally succeeded in gaining 12-month extensions on COE and HCOP funding, although by then many HCOPs had exhausted all their federal funds and lost their grantee status. The NCDHP did not succeed in restoring COE and HCOP funding to previous levels; the lame-duck 2007 Congress agreed only to maintain level funding. In FY 2008, despite a presidential veto of the Labor-HHS Appropriations bill, Congress achieved a modest increase in HCOP funding from $4 million to $10 million; still, COE funding remained unchanged at $12 million, which was sufficient to support only the 4 HBCUs. The NCDHP's goal now is to gain new funding levels that would allow all eligible institutions and disciplines to compete for COE and HCOP funding and reinvigorate national efforts to increase URM representation in the health professions.

Council on Graduate Medical Education

The Council on Graduate Medical Education (COGME), a federal think tank, was authorized by Congress in 1986 to provide an ongoing assessment of physician workforce trends, training issues, and financing policies and to recommend appropriate federal and private-sector efforts to address identified needs.

COGME is charged by its authorizing legislation to advise the Secretary of Health and Human Services and Congress on

- The supply and distribution of physicians.
- Current and future shortages or excesses of medical and surgical specialists and subspecialists.
- Issues regarding international medical graduates.
- With respect to each of the above, appropriate federal, institutional, and accrediting policies for teaching hospitals and schools of medicine and osteopathy, including financing and regulation of undergraduate and graduate medical education.
- Deficiencies in available data to conduct these analyses.

Since 1988, COGME has published 17 reports, including its 12th (1998) and 17th (2005) on minorities in medicine.[29] The 17th report reviews progress on COGME's 1998 recommendations and makes additional recommendations on

- Strengthening the pipeline to medical school.
- Strengthening upstream efforts in medical training, including admissions, medical school debt, financial assistance, and URMs in specialties and academic medicine.
- Ensuring cultural competence in medicine.

The report's appendix reviews the HCOP and COE programs, the AAMC's Project 3000 by 2000, the Health Professions Partnership Initiative, the Minority Medical Education Program, and 38 institutionally based health professions pipeline programs.

Institute of Medicine

The National Academy of Sciences Institute of Medicine is a private, nonprofit, quasigovernmental organization that conducts policy research on behalf of Congress and with the support of private foundations. The Institute of Medicine has published the following recent seminal and widely read studies relevant to URMs in health and healthcare:

- 2001: *The Right Thing To Do, the Smart Thing To Do: The Role of Diversity in the Training of Health Professionals.*[30]
- 2003: *Unequal Treatment: Confronting Racial and Ethnic Disparities in Healthcare.*[31]
- 2004: *In the Nation's Compelling Interest: Ensuring Diversity in the Health-Care Workforce.*[32]

CONCLUSION

Starting with well-established MSIs, programs to increase the representation of URMs in the health professions grew rapidly from 1987 through 2005. Fueled by federal funds, the number of COEs rose to 34, and the number of comprehensive HCOPs rose to 74. Then, the federal government cut its funding abruptly and drastically, reducing the number of COEs and HCOPs to 4 each. Several advocacy groups, supported by think tanks, have striven to restore federal funding to previous levels, so far to no avail. Meanwhile, the struggle to increase the representation of URMs in the health professions is carried on by the surviving programs, including the remaining COEs and HCOPs and new programs that, funded by state, local, and private agencies, have arisen from the ashes.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Li X, Carroll CD. *Characteristics of Minority-Serving Institutions and Minority Undergraduates Enrolled in These Institutions: Postsecondary Education Descriptive Analysis Report.* Washington, DC: National Center for Education Statistics; 2007. NCES 2008–156.
2. National Center for Research Resources. APR supplement to PHS 2590 Institutional Development Award instructions. Available at: http://www.kbrin.louisville.edu/apr/apr-instructions.pdf. Accessed September 2008.
3. Excellence in Minority Health Education and Care Act of 1987. Pub L No. 100–97.

4. US Department of Health and Human Services. *Report of the Secretary's Task Force on Black and Other Minorities Health*. Washington, DC: Government Printing Office; 1985.
5. Association of American Medical Colleges. In first joint activity, AAMC and ADEA launch new program to increase diversity of medicine and dentistry. Press release, March 2, 2005. Available at: http://www.aamc.org/newsroom/pressrel/2005/050302.htm. Accessed September 2008.
6. Testoff A, Aronoff R. The Health Careers Opportunity Program: one influence on increasing the number of minority students in schools of health professions. *Public Health Rep* 1983; 98: 284–292.
7. Wellington JS, Montero P. Equal educational opportunity programs in American medical schools. *J Med Educ* 1978; 53: 633–639.
8. Simpson CE, Aronoff R. Factors affecting the supply of minority physicians in 2000. *Public Health Rep* 1988; 103: 178–184.
9. Gwinner VM. Women's health as a model for change in academic health centers: lessons from the National Centers of Excellence in Women's Health. *J Gend Specif Med* 2000; 3: 53–55.
10. National Centers of Excellence in Women's Health homepage. Available at: http://www.4woman.gov/coe. Accessed September 2008.
11. Wong EY, Bigby J, Kleinpeter M, et al. Promoting the advancement of minority women faculty in academic medicine: the National Centers of Excellence in Women's Health. *J Womens Health Gend Based Med* 2001; 10: 541–550.
12. Collins KS. Evaluation of the National Centers of Excellence in Women's Health: sustaining the promise. *Womens Health Issues* 2002; 12: 287–290.
13. Campinha-Bacote J, Claymore-Camy D, Cora-Bramble D, et al. Transforming the face of health professions through cultural and linguistic competence education: the role of the HRSA Centers of Excellence. Available at: http://www.hrsa.gov/competence.pdf. Accessed September 2008.
14. US Office of Management and Budget. Program assessment: health professions. Available at: http://www.whitehouse.gov/omb/expectmore/detail.10000276.2002.html. Accessed August 2006.
15. US Office of Management and Budget. Detailed information on the health professions assessment. Available at: http://www.whitehouse.gov/omb/expectmore/detail.10000276.2002.html. Accessed August 2006.
16. US Government Accountability Office. *Health Professions Education Programs: Action Still Needed To Measure Impact–Report to Congressional Requesters*. Washington, DC: Government Printing Office; 2006.
17. US Government Accountability Office. Health Careers Opportunity Program: process for awarding competitive grants included independent review. Available at: http://www.gao.gov/new.items/d07137.pdf. Accessed September 2008.
18. US Department of Health and Human Services. An annotated bibliography: evaluation of pipeline development programs designed to increase diversity in the health professions. Available at: http://bhpr.hrsa.gov/healthworkforce/pipelinebibliography.htm. Accessed September 2008.
19. Saha S, Shipman SA. The rationale for diversity in the health professions: a review of the evidence. Available at: http://bhpr.hrsa.gov/healthworkforce/reports/diversity/default.htm. Accessed September 2008.
20. Patterson DG, Carline JD. Promoting minority access to health careers through health profession-public school partnerships. *Acad Med* 2006; 81(suppl): S5–S10.
21. Grumbach K, Munoz C, Coffman J, et al. Strategies for improving the diversity of the health professions. Available at: http://www.calendow.org/uploadedfiles/strategies_for_improving_the_diversity.pdf. Accessed September 2008.
22. National Institutes of Health Office of Extramural Research. NIH research supplements to promote diversity: fiscal year 2006. Available at: http://grants.nih.gov/training/fy2006_diversity_supplement_report.pdf. Accessed September 2008.
23. Terrell C. The Health Professions Partnership Initiative and working toward diversity in the health care workforce. *Acad Med* 2006 81(suppl 6): S2–S4.
24. Cohen JJ. The future of the health professions pipeline: a new call to action. *Acad Med* 2006 81(suppl 6): S60–S61.
25. National Institutes of Health Loan Repayment Programs. Clinical research repayment programs for individuals from disadvantaged backgrounds. Available at: http://www.lrp.nih.gov/about/lrp-disadv.htm. Accessed September 2008.
26. US Department of Health and Human Services. Health Disparities Research Loan Repayment Program. Available at: http://www.lrp.nih.gov/about/lrp-healthdisp.htm. Accessed September 2008.
27. Health Professions and Nursing Education Coalition. *Health Professions Programs: Over 1,000,000 Trained and Counting–FY 2008 Brochure*. Washington, DC: Association of American Medical Colleges; 2007.
28. Bednash G, North MG, Osterweis M, et al. Why health professions for diversity coalition? Why now? *Advisor* 2005; 25: 5–11.
29. Council on Graduate Medical Education. *Seventeenth Report: Minorities in Medicine: An Ethnic and Cultural Challenge for Physician Training–An Update*. Washington, DC: Department of Health and Human Services; 2005.
30. Institute of Medicine. *The Right Thing To Do, The Smart Thing To Do: The Role of Diversity in the Training of Health Professionals*. Washington, DC: National Academy Press; 2001.
31. Institute of Medicine. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: National Academy Press; 2003.
32. Institute of Medicine. *In the Nation's Compelling Interest: Ensuring Diversity in the Health-Care Workforce*. Washington, DC: National Academy Press; 2004.

Diversity in Academic Medicine No. 4

Northeast Consortium: Innovation in Minority Faculty Development

Gary C. Butts, MD,[1] Jerry Johnson, MD,[2] A. Hal Strelnick, MD,[3] Maria L. Soto-Greene, MD,[4] Beverly Williams, MS, JD,[2] and Elizabeth Lee-Rey, MD, MPH[5]

[1]Center for Multicultural and Community Affairs, Mount Sinai School of Medicine, New York, NY
[2]Center for Excellence for Diversity in Health Education and Research, University of Pennsylvania School of Medicine, Philadelphia, PA
[3]Institute for Community and Collaborative Health, Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, NY
[4]Hispanic Center of Excellence, New Jersey Medical School, Newark, NJ
[5]Albert Einstein College of Medicine, Yeshiva University, Bronx, NY

ABSTRACT

In fiscal year 2006, the US Government abruptly and drastically reduced its funding for programs to increase the racial and ethnic diversity of academic medicine, including programs to increase the development of minority medical faculty. Anticipating this reduction, 4 such programs–the Albert Einstein College of Medicine, Mount Sinai School of Medicine, University of Medicine and Dentistry in New Jersey–New Jersey Medical School, and University of Pennsylvania School of Medicine–decided to pool their resources, forming the Northeast Consortium of Minority Faculty Development. An innovation in minority faculty development, the Northeast Consortium of Minority Faculty Development has succeeded in exposing faculty trainees to research and teaching that they might not have considered otherwise, expanding the number and diversity of their mentors and role models, providing them potential access to larger and different populations and databases for purposes of research, and expanding their peer contacts. After introducing the Northeast Consortium of Minority Faculty Development, this article describes the origins and goals of each member program. ***Mt Sinai J Med 75:517–522, 2008.*** © *2008 Mount Sinai School of Medicine*

Key Words: diversity, center of excellence, minority faculty development, Northeast Consortium for Minority Faculty Development, underrepresented minority.

Address Correspondence to:

Gary C. Butts
Center for Multicultural and
Community Affairs
Mount Sinai School of Medicine
New York, NY
Email: gary.butts@mssm.edu

The US Human Resources and Services Administration (HRSA), having established Centers of Excellence (COEs) for minority education in the healthcare professions in fiscal year (FY) 1987 and providing the COEs with increasing financial support through FY 2005, abruptly and drastically reduced its support in FY 2006, forcing most COE programs to scale down or close. In FY 2005, the number of federally funded COEs was 34, and the level of US financial support was $35.6 million; in FY 2006, the number was 4, and the level of support was $12 million.

Faced with this prospect in 2005, 4 medical schools in the Northeastern United States–the Albert Einstein College of Medicine, Mount Sinai School of Medicine, University of Medicine and Dentistry in New Jersey–New Jersey Medical School, and University of Pennsylvania School of Medicine–formed a consortium to continue their pursuit of a key COE goal, namely, increasing their capacity to attract minority physician and nonphysician scientists to academic medicine and develop them into minority faculty. All 4 schools in the consortium are national leaders in minority faculty development and are based at institutions with substantial strengths and diversity of resources in research and faculty training.

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20082

© 2008 Mount Sinai School of Medicine

Called the Northeast Consortium for Minority Faculty Development (NECMFD), it focuses its efforts on 2 underrepresented minority (URM) groups, African Americans and Hispanic Americans [other URMs are Native Americans, Alaskans, and Pacific Islanders (including Hawaiians) and certain Asians (including Hmong, Cambodians, and Vietnamese)].

Working with funds from the Josiah Macy Jr. Foundation (only a fraction of the funding secured from the HRSA), NECMFD has pooled its intellectual resources to create new models of interuniversity collaboration. Members share training resources and methods, exchange trainees, and cosponsor programs, thus augmenting the strengths of all programs. In comparison with the faculty development efforts of a single institution, NECMFD provides its targeted trainees with the following unique benefits:

- Exposing the trainees to research and teaching that they might not have considered otherwise.
- Expanding the number and diversity of their mentors and role models.
- Providing them potential access to larger and different populations and databases for purposes of research.
- Expanding their peer contacts.

During its first 2 years, NECMFD enhanced the capacities of the participants by

- Conducting 2 faculty career development institutes, the first hosted by the Mount Sinai School of Medicine and the second hosted by the Albert Einstein College of Medicine.
- Meeting with the deans of the Mount Sinai and Albert Einstein medical schools to discuss ways to strengthen interuniversity collaboration and minority faculty development.
- Awarding faculty career development funds.
- Awarding pilot research funds.
- Establishing mentoring and peer networks to maximize the career development of their targeted trainees.

Proceeding in this manner, NECMFD seeks to have a significant impact on minority faculty development and become a national model that other schools can emulate.

In the meantime, each NECMFD member has also continued to build upon its own stellar URM faculty development program through a multitude of other events and programs, achievements of its URM faculty, and new grants received for health disparities research and cultural competence.

The history of each NECMFD member is described next. (For a further description of each member, see "Successful Programs in Minority Faculty Development: Overview" and "Successful Programs in Minority Faculty Development: Ingredients" elsewhere in this theme issue.)

ALBERT EINSTEIN COLLEGE OF MEDICINE AND MONTEFIORE MEDICAL CENTER

The effort to develop organizational infrastructure that explicitly supports URM faculty development at the Albert Einstein College of Medicine has been a grassroots movement that has progressed over 20 years from a single, small, marginal, hospital affiliate–based department to a COE funded by both the HRSA and National Institutes of Health (NIH) and then to a member of the privately funded consortium NECMFD.

During this gestation period, the medical school has intermittently had a part-time assistant dean for faculty development who is responsible for the entire Einstein faculty of some 3000 full-time members. The college's Office of Diversity Enhancement is led by an associate dean and an assistant dean, but their responsibilities and that of its Minority Affairs Committee have been limited to the recruitment and support of medical students.

The minority faculty development programs of the Albert Einstein College of Medicine have thus filled an organizational vacuum. Emulating fellowship programs in the Department of Family and Social Medicine (DFSM), the faculty development programs have included

- A short-lived HRSA-funded, full-time, 1-year family medicine fellowship that failed in recruiting URM fellows.
- Two specialized fellowships sponsored by the NY State Department of Health that succeeded in recruiting URM fellows: a 2-year, full-time program focused on care for adults with mental retardation and developmental disabilities and a 2-year, half-time program focused on human immunodeficiency virus/acquired immune deficiency syndrome and substance abuse.
- HRSA-funded, 2-year, part-time fellowships in academic family medicine successful in recruiting and training URM fellows as future faculty members.
- A W. K. Kellogg Foundation–funded, part-time urban health fellowship that trained interdisciplinary faculty and community leaders and activists

DOI:10.1002/MSJ

together that had a majority of URM leaders and fellows but only 2 years of funding.

- A foundation-supported, 2-year, full-time fellowship in reproductive health that included pursuing a master's degree in public health.

In 1996, the Albert Einstein College of Medicine established a clinical research training program (CRTP). The formation of the CRTP began with a meeting that the generalist departments of family medicine, internal medicine, and pediatrics had called for a different purpose, namely, to develop a common preventive medicine residency program. During the meeting, the departments learned that they all were having difficulty preparing their fellows for clinical research careers. After the departments called this to the attention of the Associate Dean for Clinical Affiliations, the latter asked interested faculty members to prepare a proposal for a formal CRTP.

Supported through the NIH K-30 program, the CRTP leads to a master's degree in clinical research methods. Requiring a 2-year, half-time commitment from each fellow's department, the CRTP provides each fellow with full tuition and stipend support (approximately 30%–40% of salary and benefits). Fellows participate in several programs administered by the Institute for Community and Collaborative Health, including core faculty meetings and mentoring, monthly conferences, bimonthly networking breakfasts, and occasional skill-building workshops. In addition, the CRTP joins the Office of Educational Resources and Assistant Dean for Faculty Development in sponsoring an annual full-day retreat for women and minority faculty held at an off-campus conference center.

As such, the CRTP serves as the infrastructure for faculty development fellowships in all of the following:

- The HRSA-funded Hispanic Center of Excellence (HCOE), which began in 2001 (5 fellows to date).
- The NIH-funded Bronx Center To Reduce and Eliminate Ethnic and Racial Health Disparities, which began in 2003 (6 fellows to date).
- The Josiah Macy Jr. Foundation–funded NECMFD, which began in 2006 (1 fellow to date).

MOUNT SINAI SCHOOL OF MEDICINE CENTER FOR MULTICULTURAL AND COMMUNITY AFFAIRS

The Center for Multicultural and Community Affairs (CMCA) was established by the Mount Sinai School of Medicine as the Office for Multicultural and Community Affairs in 1998. As the first formalized entity at Mount Sinai to increase the representation of URMs in medicine, the new office benefited from an existing network of educational pipeline programs, many of which had existed since the late 1960s, and a strong historical institutional commitment to diversity and service to poor and underserved communities.

In 2002, the CMCA became an HRSA COE for Minority Health, building on successful accomplishments with HRSA Health Careers Opportunity Program support, and began the process of formalizing and institutionalizing a systematic approach to the recruitment and development of URM faculty. In 2004, the CMCA received formal center status at Mount Sinai.

Important ingredients for the successful development of the CMCA as a COE for Minority Health include

- Support and acknowledgment from senior leadership.
- Support and direct participation in oversight, advising, and programs from department chairs and center directors.
- A data-driven process to identify disparities in racial, ethnic, and gender representation of students, housestaff, and faculty and to target interventions.
- External validation of diversity as an institutional value.
- Establishment of a centralized infrastructure to support the recruitment and development of URM faculty (the CMCA as a formal medical center has a cross-institutional charge and direct organizational and reporting relationship to the dean).
- Partnering with research training programs and projects to support the development of specific skills and research opportunities for URM faculty and trainees.

Supporting the mission of the Mount Sinai Medical Center, the CMCA aims to eliminate health disparities through a nationally recognized center structure using innovative, integrative, and coordinated approaches in the areas of community, clinical care, education, and research and to improve the health of all populations by impacting health policy and research and diversifying the healthcare workforce. The CMCA thus adds to the diversity of the Mount Sinai Medical School and Hospital and to Mount Sinai's effectiveness in serving the ethnically and racially diverse communities of East Harlem, Harlem, the Bronx, Queens, and New York City. The CMCA also serves as the interface between educational pipeline programs (including the Center for Excellence in Youth Education and the Summer Enrichment Pre-Matriculation Program, which have existed since the medical school's earliest days), institution-wide diversity initiatives, academic support for medical students, and the school's culture and medicine programs.

The key components of the CMCA Faculty Scholars Program include the following:

- The use of a faculty advisory council to identify, recruit, and support junior URM faculty.
- A focus on the development of specific skills: grant writing and scientific manuscript preparation.
- Faculty scholar participation in a research curriculum through partnering with the institutional CRTP.
- Faculty scholar group meetings and individualized coaching and use of just-in-time education to address immediate research or development needs.
- Implementation of individual faculty development plans.
- Facilitation of mentoring and networking opportunities across departments.
- Service as an advocate for negotiating within and outside departments.
- Facilitation of interest and early exposure to careers in academic medicine for URM medical students and residents.

UNIVERSITY OF PENNSYLVANIA SCHOOL OF MEDICINE: PENN CENTER OF EXCELLENCE

The University of Pennsylvania School of Medicine has a long-standing commitment to URM faculty recruitment and development.[1,2] By 1993, the school had several outstanding development programs for URM students and faculty. The school's Combined Degree Program and Physician Scholar Program had provided a variety of research experiences in biological sciences for URM members, and it eventually would have the largest number of African American and Hispanic combined-degree students among US medical schools. A substantial commitment to training URMs had been demonstrated by several of the school's centers, institutes, departments, and programs.

However, some of the school's faculty and leaders were still concerned about the lack of an adequate infrastructure to systematically recruit and develop URM faculty.[1,2] Although some departments and divisions had high proportions of URM faculty and advanced trainees (residents, fellows, and postdoctoral trainees), others had few or none. Many of the junior URM faculty were unclear about the requirements for promotion or had unrealistic expectations of how they would succeed. Others were attempting to develop clinical research without adequate training in research, without pilot funding, and without access to skilled research methodologists or statisticians. Some minority faculty members found their research impeded by multiple commitments to administrative, clinical, and teaching responsibilities.

To meet this challenge, the COE at Penn provided an infrastructure of administrative staff and faculty with expertise in program development and research methods.[1,2] The center staff developed institution-wide programs across departments and divisions for trainees who ranged from undergraduates to faculty. The result is a program with selected elements specific to the needs of investigators:

- A list of potential faculty mentors delineated according to the type of research expertise: basic science, clinical epidemiology, health services delivery, and community-based participatory research. The center shares this list with all entering faculty and holds one-on-one meetings with faculty about their choice of mentor or mentors.
- Laboratory selection. For basic science investigators, the selection of a laboratory and thus the field of basic science inquiry are concurrent. Moreover, the senior faculty member who heads that laboratory will have a profound influence on the mentee's future success in research. The center has established a short list of senior URM basic scientists to whom junior faculty (fellows and postdoctorates) are referred for advice about the various laboratories across the campus.
- Meeting with the Chair of the Committee on Appointments and Promotions. This annual meeting provides formal insight into the expectations of the Promotions Committee and the logistics of the promotions process, culminating in the delivery of a checklist created by the center that shows junior faculty what and when selected steps along the promotions pathway should be completed. Faculty

are also given a list of key resources and persons known to be helpful in career development decisions.

- Partial salary to support time devoted to research. Faculty that receive these awards are designated as URM faculty scholars and compete for these funds through a formal process that requires a structured academic plan under the supervision of a designated faculty senior faculty member.
- Research coursework. The core research course for the Center for Clinical Epidemiology MS program is made available to any Penn URM faculty at no cost. The center maintains a list of other course offerings on campus for persons with other career interests, such as the Masters in Public Health Program and the Masters in Translational Research Program. Faculty who wish to collaborate as investigators, rather than develop an independent investigative career on the model of the NIH-funded physician scientist, are offered coursework specifically designed for this level of investigator.
- Research support staff. Center staff with expertise in research methods are available for consultation in research design, assistance with establishing a database and data management system, and terminating with assistance with data analysis.

At Penn and other medical schools, all students have an introductory research experience prior to graduation.[1,2] The COE augments the opportunities of students with support for research related to health disparities. The center maintains a list of all health disparities program projects and individual initiatives on campus and matches students to investigators. For students interested in community-based participatory research, the center is able to link students with community organizations receptive to research partnerships.

UNIVERSITY OF MEDICINE AND DENTISTRY OF NEW JERSEY–NEW JERSEY SCHOOL OF MEDICINE

Established in 1991, the New Jersey School of Medicine (NJMS) HCOE has a primary goal of improving the health status of the Latino and other underserved populations by increasing the number of Latinos in the health professions and in particular medicine. As it enters its 15th year, the HCOE has focused its initiatives at enhancing the academic performance of Latinos and other individuals from other underserved backgrounds, increasing the recruitment and retention of faculty, increasing the school's diversity through the recruitment and retention of Latino faculty, and developing the capacity of all graduates to provide culturally competent healthcare services. The HCOE seeks to reduce disparities in healthcare by supporting individuals committed to advancing the goals of diversity in the medical profession and improving healthcare to underserved populations.

The programs of the HCOE are sponsored through a grant from the Division of Health Careers Diversity and Development, Bureau of Health Professions, Health Resources and Services Administration, US Department of Health and Human Services.

The diversity of our nation, combined with a shortage of individuals from disadvantaged backgrounds in the health professions, presents a significant challenge for medical education and academic medicine. Since 1972, NJMS has implemented programs in an effort to meet these challenges. With the support of the HCOE, NJMS continues to make significant strides in paving the way for student curriculum and faculty development in response to the Liaison Committee on Medical Education's standard on cultural competency. This standard maintains that medical school "faculty and students must demonstrate an understanding of the manner in which people of diverse cultures and belief systems perceive health and illness and respond to various symptoms, diseases, and treatments. Medical students should learn to recognize and appropriately address gender and cultural biases in health care delivery, while considering first the health of the patient."

To ensure that all graduates master the core competencies requisite for a culturally competent provider, a comprehensive plan that involves all 4 years of medical education was developed. In November 2003, school-wide education goals and objectives were approved. Specifically, goal 4 is "Commitment to the Health of the Community and Appreciation of Social and Cultural Diversity." Students are expected to

- Demonstrate the ability to obtain a history and physical that accommodate a patient's belief systems.
- Demonstrate cultural and linguistic competency.
- Demonstrate the ability to develop a treatment plan that accommodates the cultural, social, and economic context of the patient.

These education goals and objectives are of paramount significance, embracing the implementation of cultural competency as one of the courses without walls. The HCOE plays a critical role in developing and implementing an integrated, vertical, and longitudinal curriculum conceptualizing cultural competency and its impact in reducing health disparities in the United States.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Johnson J, Jayadevappa R, Taylor L. Extending the pipeline for minority physicians: a comprehensive program for minority faculty development. *Acad Med* 1998; 73: 237–244.
2. Johnson J, Williams B, Jayadevappa R. Mentoring program for minority faculty at the University of Pennsylvania School of Medicine. *Acad Med* 1999; 74: 376–379.

DOI:10.1002/MSJ

Diversity in Academic Medicine No. 5

Successful Programs in Minority Faculty Development: Overview

Ann-Gel S. Palermo, MPH,[1] Maria L. Soto-Greene, MD,[2] Vera S. Taylor, MSTC,[3] Ray Cornbill, MBA,[1] Jerry Johnson, MD,[4] Monica Rivera Mindt, PhD,[5,6] Desiree Byrd, PhD,[5,6] Gary C. Butts, MD,[1] Janice Herbert-Carter, MD, MGA, FACP,[7] Yvonne W. Fry-Johnson, MD, MSCR,[8] Quentin T. Smith, MD,[9] George Rust, MD,[8] and A. Hal Strelnick, MD[10]

[1] Center for Multicultural and Community Affairs, Mount Sinai School of Medicine, New York, NY
[2] Hispanic Center of Excellence, New Jersey Medical School, Newark, NJ
[3] Morehouse School of Medicine, Atlanta, GA
[4] Center for Excellence for Diversity in Health Education and Research, University of Pennsylvania School of Medicine, Philadelphia, PA
[5] Department of Psychiatry, Mount Sinai School of Medicine, New York, NY
[6] Department of Pathology, Mount Sinai School of Medicine, New York, NY
[7] Department of Medical Education, Morehouse School of Medicine, Atlanta, GA
[8] National Center for Primary Care, Morehouse School of Medicine Atlanta, GA
[9] Department of Psychiatry and Behavioral Sciences, Morehouse School of Medicine, Atlanta, GA
[10] Institute for Community and Collaborative Health, Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, NY

ABSTRACT

Despite recent drastic cutbacks in federal funding for programs to diversify academic medicine, many such programs survive and continue to set examples for others of how to successfully increase the participation of minorities underrepresented in the healthcare professions and, in particular, how to increase physician and nonphysician minority medical faculty. This article provides an overview of such programs, including those in historically black colleges and universities, minority-serving institutions, research-intensive private and public medical schools, and more primary care–oriented public medical schools. Although the models for faculty development developed by these successful schools overlap, each has unique features worthy of consideration by other schools seeking to develop programs of their own. The ingredients of success are discussed in detail in another article in this theme issue of the *Mount Sinai Journal of Medicine*, "Successful Programs in Minority Faculty Development: Ingredients of Success." ***Mt Sinai J Med 75:523–532, 2008.*** *© 2008 Mount Sinai School of Medicine*

Address Correspondence to:

Ann-Gel S. Palermo, MPH
Center for Multicultural and Community Affairs
Mount Sinai School of Medicine
New York, NY
Email: ann-gel.palermo@mssm.edu

Key Words: academic diversity, academic medicine, center of excellence, faculty development program, historically black colleges and universities, minority faculty, minority faculty development, minority-serving institution, Northeast Consortium for Minority Faculty Development, underrepresented minority.

Faculty development programs (FDPs) for physician and nonphysician scientists have been implemented in academic medical centers for decades, with proven

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20083

© 2008 Mount Sinai School of Medicine

success amply discussed in the literature.[1] However, this literature has made little or no distinction between the general development of faculty and the development of faculty members who belong to minorities underrepresented in academic medicine.[2] As defined by the Health Resources and Services Administration (HRSA), such underrepresented minorities (URMs) include African and Hispanic Americans and Native Americans, Alaskans and Pacific Islanders (Hawaiians and others), and certain Asians (Hmong, Vietnamese, and Cambodians). The development of physician and nonphysician faculty specifically belonging to such URMs is recent, and the relevant literature is accordingly sparse.

From their own perspective, URM faculty members have needs and challenges different from those of non-URM faculty members. For example, given a setting in which most minorities are historically underrepresented, a junior URM faculty member is apt to feel alone, facing unique obstacles. With relatively few URM faculty members having traveled the same path before, the junior faculty member may also lack role models and mentors, key ingredients in the success of any FDP.[3]

In order for the institution to succeed in URM faculty development, it must recognize and address the special needs of URM faculty members, providing them with the necessary support and resources, including individual attention and adequate and stable funding. Indeed, the institution must invest the same resources and care in recruiting and retaining URM faculty as it does in recruiting and retaining URM students. This investment is important, not only in its own right but also as a demonstration of the institution's commitment to the success of faculty members.

In this article, we provide an overview of successful URM faculty development, including basic principles and goals, a conceptual framework for technical skill development, and a brief description of successful institutional models. (For a more detailed discussion of this subject, see the article entitled "Ingredients of Successful Programs in Minority Faculty Development" elsewhere in this theme issue.)

BASIC PRINCIPLES AND GOALS

In order for an FDP in academic medicine to be successful, the institution and URM faculty member must first agree that the former will enable the latter to succeed, not by lowering its standards but by providing the faculty member with the necessary skills. They must also share a vision of the member's future and agree upon expectations of the member's work; for example, clinical responsibilities are only 1 aspect of the member's work.

In addition, the institution must focus on retention and promotion as well as recruitment, providing an environment that will propel the URM faculty member to remain and excel, rather than leave for private practice and other settings. More specific institutional goals are listed in Table 1.

Table 1. *Goals of the Institution in Underrepresented Minority Faculty Development.*

Environment	• Provide academically and emotionally safe environment for learning. • Assist junior faculty members in navigating the world of academic medicine. • Show faculty members that they are not alone and face the same obstacles as others. • Help junior faculty members develop a social network. • Bring faculty members, junior and senior, together. • Provide role models and mentors.
Advancement	• Guide and advance careers of junior faculty. • Help faculty member achieve professional growth in concert with the advancement of the institution. • Implement faculty development programs at different levels of the institution. • Enable underrepresented minority faculty members to advance from one level to the next, grooming them for senior faculty and/or management positions. • Enable underrepresented minority faculty members to achieve the same levels of success in publishing and grant awards as other faculty members.
Skills	• Provide skills in teaching, writing, research, and leadership. • Teach skill sets and methods necessary for advancement. • Familiarize faculty member with barriers to advancement.

CONCEPTUAL FRAMEWORK FOR DEVELOPING TECHNICAL SKILLS

Among the goals listed in Table 1 is providing the junior URM faculty member with technical skills in teaching, writing, research, and leadership. A conceptual framework for activities to deliver these technical skills is shown in Table 2. This framework is based on URM FDPs at 4 northeastern medical schools that were formerly federally funded Centers of Excellence (COEs) within the HRSA and now compose the Northeast Consortium for Minority Faculty Development (NECMFD).

The purpose of the framework is to provide leaders of academic medicine with a basis for creating or legitimizing technical skill activities for a URM FDP. As shown in Table 2, the kinds of activity are formal education; personal positioning and planning; continuous, cross-cutting, coordinated support; and monitoring and coaching. Given this framework, academic leaders can devise individual applications and activities that will provide junior URM faculty members with the skills that they need for successful advancement and promotion.

Our recommendation to use this framework comes with 3 caveats:

- When implementing a URM FDP, empowered senior leadership must consider the logistical ratio of formal and informal support structures.
- The institutional environment is like a crucible holding complex conditions and factors–some financial, others political–within which junior URM faculty members must learn to survive so that they can go on to succeed in academic medicine.
- The conceptual framework is sufficiently flexible to enable each institution to develop individualized technical skills and techniques corresponding to its needs, expectations, resources, norms, and organizational culture.

SUCCESSFUL INSTITUTIONAL MODELS

Despite recent drastic cutbacks in federal funding for programs to diversify academic medicine, several such programs survive and continue to set examples for others of how to successfully increase the participation of minorities underrepresented in the healthcare professions.

Among the surviving diversity programs are those of minority-serving institutions (MSIs) and programs that previously received federal funds as COEs from the HRSA. The MSIs include Howard University School of Medicine and Morehouse Medical College, both serving African Americans; the Sophie Davis School of Medicine of the City University of New York, serving African and Hispanic Americans; and the University of Washington, serving Native Americans. The former COEs include the medical schools of the University of California at San Diego, University of Minnesota–Duluth, University of Pennsylvania, and University of Medicine and Dentistry of New Jersey, as well as the Albert Einstein College of Medicine and Mount Sinai School of Medicine. Following the loss of federal funding and designation as HRSA COEs, the last 4 schools formed NECMFD.

The successes of these schools and the models that they follow are described next.

Table 2. *Conceptual Framework for Delivering Technical Skill Development Activities to Underrepresented Minority Faculty Members.*

Kind of Activity	Description
Formal education	Specialized training and education programs, each usually institutionally sponsored and offered as a master's level program.
Personal positioning and planning	Process of examining, assessing, identifying, and implementing strategies leading toward their fulfillment within the academic medicine environment.
Continuous, cross-cutting, coordinated support	Many forms cutting across all realms of responsibility, including teaching, education, community and institutional service, and personal lifestyle. Support must be long-term, providing a continuous pathway from a junior rank to a senior rank, and must be coordinated with resources, time, and attention (ie, salary support for protected time to focus on career development).
Mentoring and coaching	Working and active partnership between the junior underrepresented minority faculty member and seasoned faculty member(s) whose knowledge and experience of institutional culture and environment are sufficiently deep and rich to enable the junior faculty member to accomplish specific goals.

Albert Einstein College of Medicine/Montefiore Medical Center Institute for Community and Collaborative Health

Formerly an HRSA COE and now a member of NECMFD, the Albert Einstein Montefiore Institute for Community and Collaborative Health (ICCH) was founded by the dean of the medical college in 1999. Its mission is to improve the health of the Bronx community through research, education, service, and collaborative partnerships and more specifically to

- Support education, service, and research in community and preventive medicine.
- Support interdisciplinary and interinstitutional collaborations that promote community health.
- Promote and support academic-community partnerships in education, service, and research.
- Assist the Albert Einstein College of Medicine in living up to its name as an institution of science and social responsibility that fosters a culture of service and community citizenship.

The ICCH began by consolidating several URM FDPs that had grown separately over the previous 20 years, including an HRSA fellowship program, a clinical research training program, and the National Institutes of Health (NIH)–funded Bronx Center To Reduce and Eliminate Ethnic and Racial Health Disparities (Bronx CREED).

Both the initial and subsequent ICCH FDPs have been coordinated with a part-time assistant dean for faculty development. Although this position is responsible for the development of approximately 3000 full-time faculty dispersed among 3 boroughs and 2 suburban counties, it does not have a discrete budget, and the position has more often been vacant than filled.

Nevertheless, the ICCH is now the administrative home for almost $3 million in annual funding from state, federal, and private sources as well as direct budget support from the dean of the medical college. In addition to administering funded grant programs, it provides tuition reimbursement for URM faculty to pursue master's and doctoral degrees in public health to advance their careers. It also provides individualized mentoring, letters of recommendation, guidance, and technical assistance (eg, proposal editing) for career development awards, including the Robert Wood Johnson Foundation's Harold Amos Minority Faculty Development Awards, the NIH K-07 Mentored-Clinical Research Career Development Awards, and a National Heart, Lung, and Blood Institute (NHLBI) Health Disparities and Cultural Competency Academic (Curriculum Development) K-07 Award.

Through its numerous grant awards and community-academic partnership, the ICCH also provides leadership opportunities for URM faculty as coprincipal investigators, coinvestigators, codirectors, and core directors and funding and cosponsorship for the institution-wide Albert Einstein College of Medicine Women and Minority Faculty Career Day, an annual faculty development retreat begun in 2004. In addition, the ICCH advocates on behalf of URM faculty for inclusion in such important research projects as the NHLBI's Hispanic Community Health Study (now known as Study of Latinos), a Framingham-like longitudinal cohort study that has 1 of its 4 sites at Einstein in the Bronx. Through its community partnerships, the ICCH recruits community-based faculty to take on important roles in building community-academic partnerships. These include the Medical Director of the Einstein Community Health Outreach Free Clinic, who has joined the core faculty of the Hispanic Center of Excellence (HCOE), and the Executive Director of the Bronx Health Link, who will be leading an NIH-funded community research review board.

ICCH's community-academic partnerships, especially Bronx CREED, have also played the leading role in developing the community engagement core required in Einstein's successful application for the NIH Roadmap initiative, the Clinical and Translational Science Award. Each iteration of Einstein's application strengthened this component, which was already judged "outstanding" in its first review. The proposal's community engagement core director was a Latina graduate of the HCOE faculty development fellowship.

As shown by the HCOE, federal grant funds from agencies such as the HRSA are vulnerable to changing political winds. Although institutional support can bridge some of the effects of dramatic funding cuts, Einstein's HCOE faculty development fellowship could not be continued. Similarly, tight funding at NIH led a reduction of the 2-year-long Bronx CREED faculty development fellowships from 2 per year to 1 every other year.

Financial issues that often plague earlier stages in the health career pipeline also challenge minority faculty. With considerable debt burdens and family responsibilities, faculty of color often feel pressure to moonlight or leave academic medicine altogether for the higher incomes of private practice. They may not have the financial options to extend their training beyond residency as research or faculty development fellows if stipends are not competitive with their other career options.

In summary, through trial and error, the Albert Einstein Medical College and Montefiore

Medical Center have developed a new organizational infrastructure to support URM faculty development, using a variety of grant-funded revenue streams and providing advocacy and a variety of opportunities for junior and senior URM faculty to advance their careers as leaders, educators, investigators, and role models.

Necessity being the mother of invention, the Einstein/Montefiore FDP has demonstrated that such an organizational infrastructure can be grown through grassroots efforts, successful grantsmanship, and committed senior faculty and staff. In the near future, it will realize an invaluable return on its investment in increasing senior URM faculty from among its own students, residents, fellows, and junior faculty. Through their words and actions, these future URM faculty leaders will provide Albert Einstein Medical College with the most convincing argument for using its own resources to institutionalize the FDP organizational infrastructure and create a culture that recognizes and values diversity, retains and rewards talent, and ensures that URM faculty thrive.

Howard University School of Medicine

One of the historically black colleges and universities (HBCUs), Howard University presently consists of 12 colleges and institutions of higher learning, including associated colleges of medicine, dentistry, pharmacy, nursing, and law.[4] As a comprehensive, research-oriented HBCU, Howard University provides an educational experience of exceptional quality to students of high academic potential, with a particular emphasis on the provision of educational opportunities to promising black students.

The university is also dedicated to attracting and sustaining a cadre of faculty who are, through their teaching and research, committed to the development of distinguished and compassionate graduates and to the quest for solutions to human and social problems in the United States and throughout the world.[4]

To this end, the university has established the Center for Excellence in Teaching, Learning, and Assessment (CETLA). Funded by the HRSA, CETLA has 3 goals[4]:

- Empower the faculty to teach more effectively, especially through the reflective use of technology.
- Achieve measurable improvement in students' learning.
- Provide national leadership in the movement to strengthen college teaching.

To fulfill these goals, CETLA conducts the following multifaceted activities[4]:

- Training, which consists of workshops, guest speakers, and a summer institute, in which one may participate in a 3-day course to become certified as a blackboard teacher and receive certification as a distance-education instructor by enrolling in a week-long distance-learning topical online course.
- Support, which consists of open labs, with on-duty staffing for assistance. User group meetings include face-to-face opportunities for trouble shooting and strategy sharing between individuals experimenting with similar teaching methods such as distance learning or across-the-curriculum writing. Conversely, there are online forums for shared mentoring, troubleshooting, and communication with peers finding themselves in similar situations. Private consultations are offered as well as special services for faculty with disabilities.
- Research, in which CETLA assists faculty with educational research in its teaching laboratories.[4]

Morehouse School of Medicine

Among formally designated MSIs, only the Morehouse School of Medicine has an HRSA-funded FDP.[5] In 1992, Morehouse started the FDP with the goal of increasing the nation's URM faculty in family medicine. In 2001, the program was expanded to include all primary care specialties, and in 2002, it established an executive FDP to include participants from across the nation.

The FDP curriculum involves 35 workshops divided into topical areas, each crucial to the successful development of medical faculty.[5,6] Initially longitudinal, the workshops were administered over 4 quarters beginning in September and ending in June. Today, the program is modular, with each topical area presented in one-half day workshops over a period of 4 to 6 weeks. A participant can complete 1 or more areas, each associated with continuing medical education credit. Success is measured by attendance, active participation, and the completion of assignments and all required projects. Participants who successfully complete all topical areas are designated program completers (Figure 1).

To increase the pool of URM academic teachers and researchers, the Morehouse FDP includes community-based preceptors in its target audience. Some community preceptors have not previously considered the possibility of becoming academic clinicians. Through guided development, the participants produce deliverables that are actual end products, such as letters to the editor, case reports, and completed and submitted grant applications. Thus, Morehouse demonstrates to its participants who are community preceptors not yet directly involved in

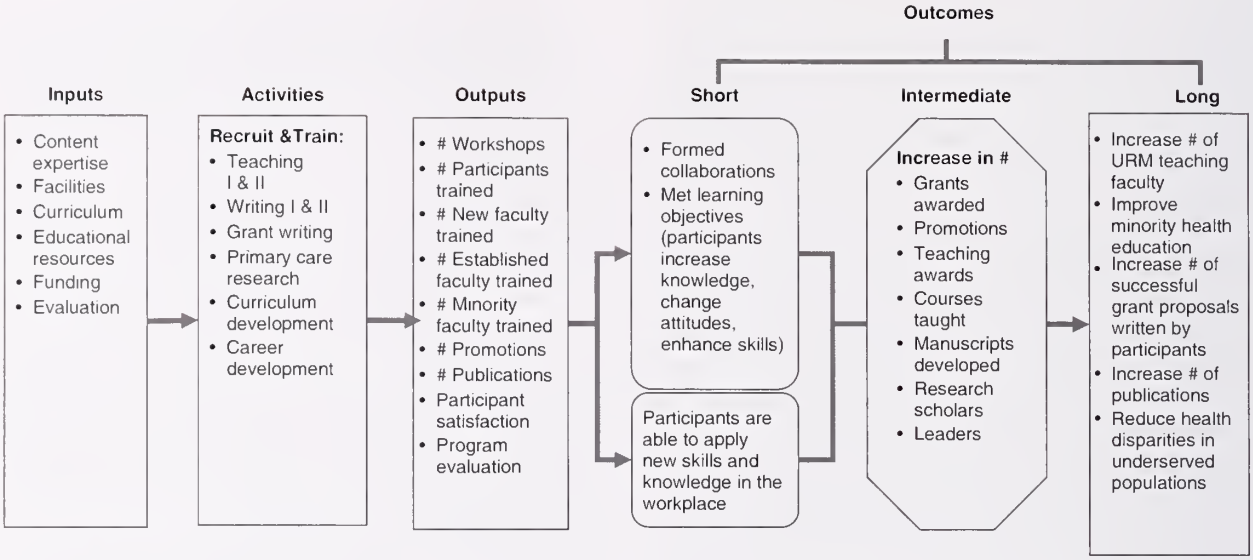

Outcomes
Inputs
Activities
Outputs
Short
Intermediate
Long
Content expertise
Facilities
Curriculum
Educational resources
Funding
Evaluation
Recruit &Train:
Teaching I & II
Writing I & II
Grant writing
Primary care research
Curriculum development
Career development
# Workshops
# Participants trained
# New faculty trained
# Established faculty trained
# Minority faculty trained
# Promotions
# Publications
Participant satisfaction
Program evaluation
Formed collaborations
Met learning objectives (participants increase knowledge, change attitudes, enhance skills)
Participants are able to apply new skills and knowledge in the workplace
Increase in #
Grants awarded
Promotions
Teaching awards
Courses taught
Manuscripts developed
Research scholars
Leaders
Increase # of URM teaching faculty
Improve minority health education
Increase # of successful grant proposals written by participants
Increase # of publications
Reduce health disparities in underserved populations

Fig 1. Morehouse School of Medicine faculty development training model. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

the formal academic pathway that these are indeed competencies that they can master. Some of these preceptors go on to become adjunct faculty and/or continue to teach in residency and undergraduate medical education programs. In addition, some chief residents and URM medical students participate in the academic medicine elective, thus contributing to the pipeline for future URM academic physicians.

After completing the Morehouse FDP, faculty can further their academic careers by entering the Masters of Science in Clinical Research Program, which is funded by the NIH.[6] This program has been completed by several Morehouse FDP participants.

The Morehouse FDP has proven to be an effective model for developing URM clinician educators, research-investigator scholars, and leaders in the healthcare professions.[6] To date, the executive and modular (longitudinal) programs have been completed by 197 participants, of whom 88% have been black/African American. Most currently hold faculty rank and are actively involved in teaching at US medical schools or residency training programs. Demonstrating their academic productivity, the participants had been credited with a total of 195 publications by 2007 (Table 3). (Before this, most participants had no prior experience with publishing except for sending a letter to the editor in MSM Writing 101.) The participants also reported 184 presentations and 24 grants, including a $1.4 million grant to enable high-school and college students to pursue careers in health professions (the participant's first grant award) and, in collaboration with another organization, a $100,000 Pathways to Health Professions Grant to mentor 100 high-school students through a 7-week pediatric mini–medical school (which also introduced students to healthcare careers other than medicine). Finally, the participants reported 9 promotions over a period of 5 years. One of the charter graduates became a valued MSM faculty member who won teaching awards and served as Medical Director of the MSM Family Practice Center. Another advanced to the position of chief medical officer for an HBCU hospital, and a third was recently promoted to a full MSM professorship and is now a candidate for a department chair. Still others serve as residency program directors.

Table 3. *Morehouse School of Medicine Faculty Development Publications.*

Type	Number
Abstracts	4
Articles	67
Book reviews	7
Case reports	16
Chapters	12
Editorials	3
Grants	24
Letter to the editor	62
Total	195

As shown in Table 4, participants rate the program highly. The strengths they cite most often are small class size, interactivity, excellent instructors, practical hands-on experience, and the integration of theory and practice.

Table 4. *Top Ten Strengths of the Morehouse School of Medicine Faculty Development Program According to Its Participants.*

1. You demystified the process of grant writing.
2. Learned to give a large amount of information in a short time.
3. I've been inspired to take on projects that before seemed like a dream.
4. This is an excellent series that helped me hone my clinical skills.
5. Enthusiastic instructors.
6. Provision of multiple resources.
7. It was enlightening because it allowed me to experience analyzing data.
8. Simplified research concepts so we could grasp them.
9. Will recommend it [curriculum development] to my colleagues.
10. Presenting a talk in the auditorium simulated a real-life situation.

Table 5. *Morehouse School of Medicine Faculty Development Participant Demographic Data.*

Racial/Ethnic Group	Males	Females	Total
American Indian/Alaska native	0	1	1
Asian	3	9	12
Black/African American	68	106	174
Hispanic/Latino	1	0	1
Native Hawaiian/other Pacific Islanders	0	0	0
White	7	2	9
Total	79	118	197

NOTE: This table was adapted from Rust *et al.*[6]

Mount Sinai School of Medicine Center for Multicultural and Community Affairs

Responding to the need for more URM medical faculty, the approach of the Mount Sinai Center for Multicultural and Community Affairs (CMCA) has been comprehensive and strategic, building on lessons learned from the literature and experiences of other URM FDPs. These include

- Increasing awareness, interest, and preparation for URM medical students and residents for academic careers (to support a faculty pipeline).
- Incentivizing potential applicants and department chairs.
- Raising awareness of the value of diversity to department chairs.
- Supporting rigorous development of URM faculty once recruited.

The CMCA is centrally positioned within the Mount Sinai School of Medicine organizational structure (Figure 2). The CMCA director, an associate dean, reports directly to the dean of the medical school. Recently, the title of the CMCA director was named the Associate Dean for Diversity Programs, Policy, and Community Affairs. The new position involves expanded responsbilities, including the management of diversity programs at all levels of education and training, an increase in service and educational opportunities in the surrounding communities, and an increase in minority and health disparities research.

Programmatically, the CMCA focuses on

- Community engagement and service learning.
- Culture and health curriculum.
- Academic support and student counseling.
- Medical student research.
- URM recruitment and retention (medical school and postgraduate training).
- URM student professional development.
- Advising medical student organizations and activities.
- Educational pipeline programs.
- URM faculty recruitment and development.
- Medical education and research policy.

Since 2002, the CMCA has recruited 16 URM clinical and basic science faculty members and promoted 12 basic science and clinical URM faculty members in rank. From 2001 (when the CMCA was established) to 2007, the URM faculty grew from 3% to 7%. During the period of 2004–2007, the CMCA engaged 25 junior faculty members, including 5 from North General Hospital, in formal research training and academic development programs and supported 48 predoctoral students in mentored research.

Sophie Davis School of Biomedical Education of City University of New York

Serving a large proportion of African and Hispanic Americans among its students, the Sophie Davis School of Biomedical Education has been designated an MSI. Although it does not have a faculty development office, its program does include faculty development on a limited basis. Its initiatives fall into 3 categories:

- Intermittent planned sessions such as review examinations from the National Board of Medical Examiners and the National Board of Medical Examiners question writing workshop.
- A weekly slide-lecture series from the International Association of Medical Science Educators. The

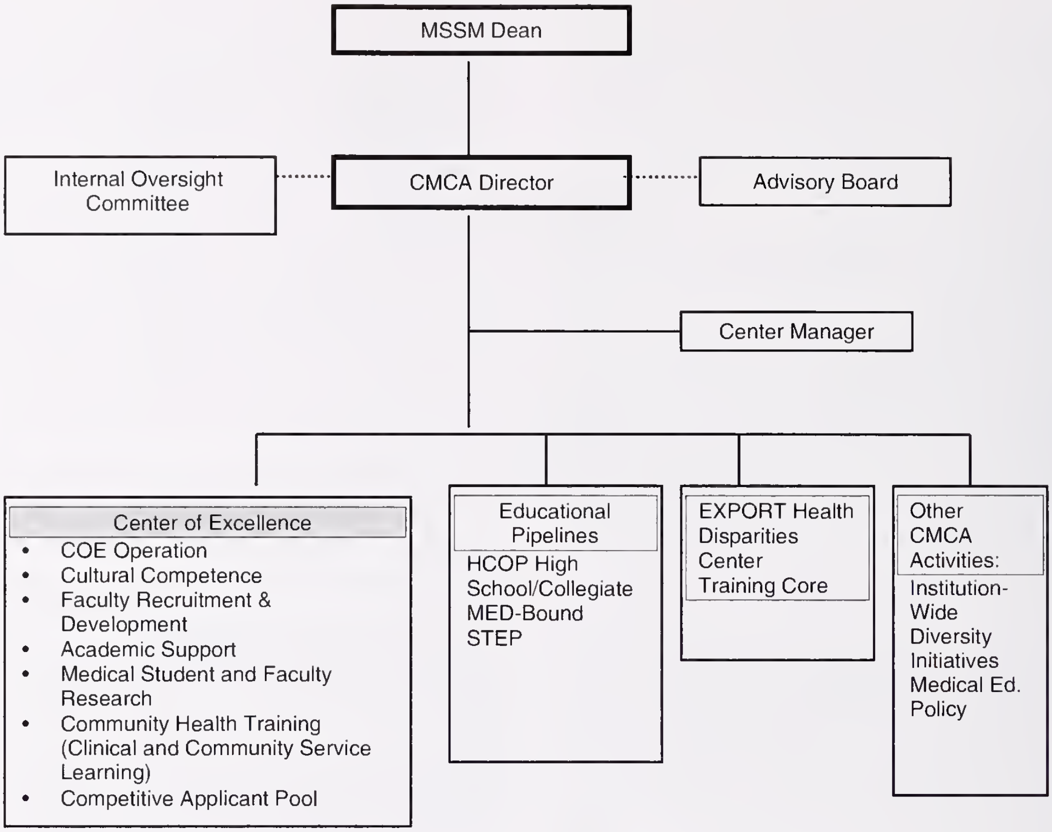

MSSM Dean
Internal Oversight Committee
CMCA Director
Advisory Board
Center Manager
Center of Excellence
• COE Operation
• Cultural Competence
• Faculty Recruitment & Development
• Academic Support
• Medical Student and Faculty Research
• Community Health Training (Clinical and Community Service Learning)
• Competitive Applicant Pool
Educational Pipelines
HCOP High School/Collegiate
MED-Bound
STEP
EXPORT Health Disparities Center Training Core
Other CMCA Activities:
Institution-Wide Diversity Initiatives
Medical Ed. Policy

Fig 2. Mount Sinai CMCA: organizational structure. **Abbreviations:** CMCA, Center for Multicultural and Community Affairs; COE, Center of Excellence; EXPORT, Excellence in Partnerships for Community Outreach, Research, MED-Bound, Medicine Bound; STEP, Science Technology Entry Program and Training; HCOP, Health Careers Opportunity Program; MSSM, Mount Sinai School of Medicine.

slides are webcast, and the lecture is delivered via conference phone. One recent topic was the development of leadership skills.

- A public seminar series presented 3 to 4 times per semester, involving presentations by the faculty on their research or by outside speakers on cutting-edge topics. One recent theme was translational research.

University of California at San Diego School of Medicine

The University of California at San Diego HCOE Scholars Program has 77 alumni. Of 58 scholars who have completed medical training, 6 (10%) are now salaried faculty in academic clinical and administrative positions (unpublished data).

University of Minnesota Center of American Indian and Minority Health

Since 2004, the University of Minnesota's Center of American Indian and Minority Health has hosted a series of informational programs for Native American medical students and residents from across the country. The purpose of the multiyear program is to expose participants to the utility of research in Native American communities. The seminars educate the participants on the means by which research can be accomplished in a manner consistent with traditional Native American values and within the context of a Native American worldview. By first educating participants about the needs of Native American communities and the mechanisms that research employs in addressing those needs, the Center of American Indian and Minority Health has been successful in showing a meaningful purpose for research involving Native American communities and the positive role of the investigator in that process. During the first 3 years, community collaboration and individualized mentorship with a Native American researcher resulted in the completion of several medical student research projects.

University of Pennsylvania School of Medicine Penn Center of Excellence

The overall goal of the Penn COE at the University of Pennsylvania School of Medicine is to establish, facilitate, and conduct programs that will enhance the health of URMs, particularly African Americans and Hispanics, by training URM medical students and medical faculty for leadership positions in the health professions workforce. Initiated in 1993, the center's programs for URM faculty development, cross-cultural medical education, and minority health research have become models for other institutions both in the region and across the nation.

Although mentoring is critical to faculty development, the University of Pennsylvania School of Medicine has recognized the need for a comprehensive and interconnected FDP. The Penn FDP thus encompasses 4 levels of trainees: premedical students, medical students, advanced trainees (residents, fellows, and postdoctoral trainees), and faculty. The principal student-recruitment component is the undergraduate premedicine enrichment program. The medical student component provides general counseling, research development, and activities to enhance performance in the clinical courses. The components for advance-degree trainees and faculty consist of training in research methods, mentoring, teaching skills, and scientific writing skills as detailed in an article by Johnson and colleagues.

Additional programs pertinent to URM faculty development are the following:

- A unique longitudinal secondary training program in 2 area high schools (one predominantly African American and the other predominantly Hispanic).
- A comprehensive faculty recruitment and development program responsible for an increase in the number of minority faculty from 27 to 119.
- Curricula related to minority health and cultural competence involving the entire medical student class.
- Campus-wide minority health disparities research programs that rival those of any academic medical center in the United States.

University of Washington School of Medicine

Designated an MSI because of its significant proportion of Native American students, the University of Washington School of Medicine is also the site for a Native American Center of Excellence (NACOE). Founded in 1992, the NACOE was initially funded by the HRSA and is now funded by the University itself.

The NACOE's primary goals are to

- Attract Native American and Alaskan native students into the study and practice of medicine.
- Recognize Native American healing traditions and integrate them into the Western medical education that the medical students pursue.
- Provide an Indian Health Pathway certification program.
- Provide continuing professional development of Native American physicians.
- Encourage Native American researchers and research into Native American health issues.
- Support multicultural outreach and diversity efforts at the University of Washington with and through the School of Medicine's Office of Multicultural Affairs activities and programs.

In addition, the NACOE has initiated a program to develop Native American medical faculty. The NACOE faculty development Web site offers online videos of lectures that both further faculty development and provide healthcare information of general interest to the Native American community. The topics are ones regularly overlooked or often given superficial treatment; examples include "Health Prevention for Native Americans," "An Evidence-Based Approach to Herbals and Supplements," "Improving Feedback for Everyone," and "Unequal Treatment: Confronting Health Disparities in Native Americans."

SUMMARY

Despite recent drastic cutbacks in federal funding for programs to diversify academic medicine, several such programs survive and continue to set examples for others of how to successfully increase the participation of minorities underrepresented in the healthcare professions and, in particular, how to increase physician and nonphysician minority medical faculty.

Successful minority medical FDPs have been developed in a variety of academic settings, including HBCUs (eg, Howard University and the Morehouse School of Medicine), MSIs (eg, Sophie Davis School of Biomedical Education), research-intensive private and public medical schools (eg, Albert Einstein College of Medicine, Mount Sinai School of Medicine, and the medical schools of the Universities of Pennsylvania, Washington, and California at San Diego), and more primary care–oriented public medical schools (eg, University of Minnesota–Duluth and University of Medicine and Dentistry of New Jersey–New Jersey Medical School).

Although the models for faculty development developed by these successful schools overlap, each has unique features worthy of consideration by other schools seeking to develop programs of their own. The ingredients of success are discussed in detail in another article in this theme issue, "Successful Programs in Minority Faculty Development: Ingredients of Success."

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Harris DL, Krause KC, Parish DC, Smith MU. Academic competencies in medical faculty. *Fam Med* 2007; 39: 343–350.
2. Soto-Greene ML. Minority faculty development. Paper presented at: 2004 Meeting of the American Association of Medical Colleges; 2004; Philadelphia, PA.
3. Price EG, Gozu A, Kern DE, et al. The role of cultural diversity climate in recruitment, promotion, and retention of faculty in academic medicine. *J Gen Intern Med* 2005; 20: 565–571.
4. Howard University. Center for Excellence in Teaching, Learning, and Assessment. Available at: http://www.cetla.howard.edu/about/mission.html. Accessed September 2008.
5. Rust GS, Taylor VT, Morrow R, Everett J. The Morehouse Family Practice Faculty Development Program–methods and three-year outcomes. *Fam Med* 1998; 30: 162–167.
6. Rust G, Taylor V, Herbert-Carter J, et al. The Morehouse Faculty Development Program–evolving methods and 10-year outcomes. *Fam Med* 2006; 38: 43–49.
7. Howard University. Message from the president. Available at: http://www.founders.howard.edu/president-reports/mission.htm. Accessed September 2008.

DOI:10.1002/MSJ

Mount Sinai Journal of Medicine 75:533–551, 2008

Diversity in Academic Medicine No. 6

Successful Programs in Minority Faculty Development: Ingredients of Success

Sandra P. Daley, MD,[1] Ann-Gel Palermo, MPH,[2] Marc Nivet, EdD,[3] Maria L. Soto-Greene, MD,[4] Vera S. Taylor, MSTC,[5] Gary C. Butts, MD,[2] Jerry Johnson, MD,[6] A. Hal Strelnick, MD,[7] Elizabeth Lee-Rey, MD, MPH,[8] Beverly Williams, MS, JD,[6] Jocelyn Dorscher, MD,[9] Wanda D. Lipscomb, PhD[10] Sherria McDowell, MPH,[2] Ray Cornbill, MBA,[2] Monica Rivera Mindt, PhD,[11,12] Janice Herbert-Carter, MD, FACP,[13] Yvonne W. Fry-Johnson, MD, MSCR,[14] Quentin T. Smith, MD,[15] George Rust, MD,[14] and Kofi Kondwani, MD[14]

[1] Division of Community Pediatrics, University of California at San Diego, La Jolla, CA
[2] Center for Multicultural and Community Affairs, Mount Sinai School of Medicine, New York, NY
[3] Josiah Macy Jr. Foundation, New York, NY
[4] Hispanic Center of Excellence, New Jersey Medical School, Newark, NJ
[5] Morehouse School of Medicine, Atlanta, GA
[6] Center for Excellence for Diversity in Health Education and Research, University of Pennsylvania School of Medicine, Philadelphia, PA
[7] Institute for Community and Collaborative Health, Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, NY
[8] Albert Einstein Medical College, Yeshiva University, Bronx, NY
[9] Center of American Indian and Minority Health, University of Minnesota Medical School, Duluth, MN
[10] Center of Excellence for Diversity in Medicine, Michigan State University, East Lansing, MI
[11] Department of Psychiatry, Mount Sinai School of Medicine, New York, NY
[12] Department of Pathology, Mount Sinai School of Medicine, New York, NY
[13] Department of Medical Education, Morehouse School of Medicine, Atlanta, GA
[14] National Center for Primary Care, Morehouse School of Medicine, Atlanta, GA
[15] Department of Psychiatry and Behavioral Sciences, Morehouse School of Medicine, Atlanta, GA

ABSTRACT

This article describes the ingredients of successful programs for the development of minority faculty in academic medicine. Although stung by recent cuts in federal funding, minority faculty development programs now stand as models for medical schools that are eager to join the 140-year-old quest for diversity in academic medicine. In this article, the ingredients of these successful faculty development programs are discussed by experts in minority faculty development and illustrated by institutional examples. Included are descriptions of program goals and content, mentoring and coaching, selecting participants, providing a conducive environment, managing the program, and sustaining support. This article is a companion to another article, "Successful Programs in Minority Faculty Development: Overview," in this issue of the *Mount Sinai Journal of Medicine*. **Mt Sinai J Med 75:533–551, 2008.** © *2008 Mount Sinai School of Medicine*

Address Correspondence to:

Sandra P. Daley
Division of
Community Pediatrics
University of California at
San Diego
iLa Jolla, CA
Email: sdaley@ucsd.edu

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20084

© 2008 Mount Sinai School of Medicine

Key Words: diversity, academic medicine, center of excellence, faculty development program, minority faculty, minority faculty development, minority-serving institution, Northeast Consortium for Minority Faculty Development, underrepresented minority.

This article concerns the successful design and implementation of programs to develop medical faculty drawn from underrepresented minorities (URMs): African and Hispanic Americans; Native Americans, Alaskans, and Pacific Islanders, including Hawaiians; and certain Asians, including Hmong, Vietnamese, and Cambodians. In some respects, the needs of URM and non-URM faculty are the same, and they require the same program ingredients, such as an emphasis on skills training and mentoring by senior faculty. In other respects, the needs of URM faculty are distinct and require a particular response, such as help with finding mentors, cultural matching, nurturing teachers, and funds to buy out time required for development.

Through years of experimentation, practice, evaluation, and modification, URM faculty development programs (FDPs) have identified the ingredients necessary for their success. These ingredients can be subsumed under the following headings:

- Setting program goals and content.
- Mentoring and coaching.
- Selecting participants.
- Providing a conducive environment.
- Managing the program.
- Sustaining support.

The discussion of each ingredient in this article includes examples drawn from the successful programs shown in Table 1. (For an introduction to most of these programs, see the article entitled "Successful Minority Faculty Development Programs: Overview" elsewhere in this theme issue.)

SETTING PROGRAM GOALS AND CONTENT

Technical Skills Training

A fundamental ingredient of any successful FDP, minority or otherwise, is training in a broad set of technical skills. Based on the experience of the Northeast Consortium for Minority Faculty Development (NECMFD), the required skills are those shown in Table 2.

Examples from specific programs are described next.

Morehouse School of Medicine

The Morehouse modular and executive programs include Teaching 101–102, Writing for the Medical Literature 101–102, Grant Proposal Writing, Health Services Research, Leadership Skills, Career Development, and Curriculum Development. During these programs, participants gain skills in

- Small-group and large-group teaching.
- Writing objectives.
- Effectively addressing the needs of various learning styles.
- Using projected and nonprojected media.
- Analysis of large Medicare/Medicaid data sets.
- Preparing and presenting a poster.
- Writing and submitting both a letter to the editor and a case report.
- Time management.
- Networking.
- Negotiating skills.
- Development of a curriculum for a 30-day clerkship.
- Computer skills.
- Cultural competence.

All sessions are taught in a highly interactive manner with hands-on learning in the computer laboratory, role playing in the classroom, and opportunities for presentation to other members of the medical school community.

University of Pennsylvania School of Medicine–Penn Center of Excellence

The Penn Center of Excellence (COE) recognizes that some URM faculty need to augment their skills in teaching and in writing grants and manuscripts. Writing skills are critical to a faculty member's development as a clinical scholar or investigator.[1,2]

Consequently, the COE has established annual workshops in teaching, grant writing, and manuscript writing, all led by experts from within and outside the university.[1,2] A typical workshop is limited to the top 10 to 15 faculty members and is conducted over a 1- to 3-day period.

Preparing for Research

In order to understand, conduct, and manage clinical and biomedical research activities, a junior URM faculty member may require formal education in clinical and basic science research provided in a master's level program such as the Clinical Research Training Program (CRTP). Funded by the National Institutes of Health (NIH) through the Clinical Research Curriculum Award, the CRTP is designed

DOI:10.1002/MSJ

Table 1. *Medical Schools with Successful Programs in Minority Faculty Development.*

Institution	Location
Albert Einstein College of Medicine/Montefiore Medical Center Institute for Community and Collaborative Health*	Bronx, NY
Howard University School of Medicine	Washington, DC
Michigan State University Center of Excellence for Cultural Diversity in Medical Education	East Lansing, MI
Morehouse School of Medicine	Atlanta, GA
Mount Sinai School of Medicine Center for Multicultural and Community Affairs*	New York, NY
Sophie Davis School of Medicine of City University of New York	New York, NY
University of Arkansas College of Medicine	Little Rock, AR
University of California at San Diego School of Medicine Hispanic Center of Excellence	San Diego, CA
University of Medicine and Dentistry of New Jersey–New Jersey Medical School*	Newark, NJ
University of Minnesota Center of American Indian and Minority Health	Twin Cities, MN
University of Pennsylvania School of Medicine–Penn Center of Excellence*	Philadelphia, PA
University of Washington School of Medicine Native American Center of Excellence	Seattle, WA

* Member of the Northeast Consortium of Minority Faculty Development.

to attract talented individuals to clinical research and provide them with the skills needed to translate basic discoveries into clinical treatments.[3] Participating institutions are expected to provide high-quality, multidisciplinary, didactic training, including core courses in the fundamental scientific and statistical skills, methodologies, and theories necessary for the well-trained, independent, clinical researcher.

Mount Sinai School of Medicine Center for Multicultural and Community Affairs

In order to complement its FDP with a formal 2-year clinical research curriculum, the Mount Sinai Center for Multicultural and Community Affairs (CMCA) has formed partnerships with the Mount Sinai CRTP and Masters of Public Health Program. The first year of the curriculum is devoted to didactic training, and the second year is primarily devoted to research under the guidance of program advisors and the research mentor of each trainee.

The CMCA also has research collaborations with numerous other Mount Sinai Medical Center departments and East Harlem community entities including: the Excellence in Partnerships for Community Outreach, Research, and Training Center (Department of Health Policy), Excellence Centers To Eliminate Ethnic/Racial Disparities (Department of Health Policy), East Harlem Diabetes COE (Department of Health Policy), East Harlem Partnership for Diabetes Prevention, Growing Up Healthy in East Harlem, and Neuro-AIDS Training Institute. Moreover, the CMCA collaborates in the development of NIH training grants with the Mount Sinai Departments of Pediatrics, Psychiatry, Internal Medicine (Renal), Emergency Medicine, and Genetics.

In developing the research skills of URM faculty, the CMCA uses an approach called *individualized just-in-time education.* While participating in CRTP research coursework, junior URM faculty members also participate in research and may at times need a research skill before the coursework provides it to them. To solve this problem in timing, the CMCA has engaged expert consultants in research methods to work one-on-one with junior CRTP program participants at various stages of their research. Supplementing the mentors in clinical and basic science research, the consultants provide tutorials on research topics, methodology, and analysis as well as overall direction, including navigation toward more junior-level projects as appropriate.

University of California at San Diego School of Medicine

As part of the FDP at the Hispanic Center of Excellence (HCOE) of the University of California at San Diego, medical students enrolled in the HCOE Scholars Program receive guidance from senior investigators, engage in structured research experiences, and have the opportunity to complete research pilot projects that can ultimately lead to funding opportunities. There are 3 major categories of research experiences: basic science research, clinical research, and applied research in community settings. Scholars may also engage in health promotion and the evaluation of programs designed to improve health literacy and health indicators in underserved communities. HCOE scholars are encouraged to establish collaborations with community partners that include converting their activities into research projects that benefit both the community and the university.

By the second year of medical school, each HCOE scholar selects a research mentor, an

Table 2. *General Technical Skill Development Areas Requiring Adequate Proficiency.*

Scholarship and research
- Research proposals and portfolio
- Basic concepts in epidemiology and biostatistics
  - Statistics for the nonstatistician
- Developing outcome and clinical research
- Evaluating research proposals
- Institutional review board and ethics 101
- Working with nontraditional research partners (ie, community partners)
- How to be a peer reviewer and remain a colleague
- Data analysis
- Grantsmanship, including scientific writing and research grant writings skills

Communication styles and skills
- Communication and leadership tools
- Communication: getting it in writing
- Cross-cultural communication competencies
- Presenting work
- Working with the media
- Videotape assessment
- Presenting work orally and visually
- Negotiation basics: practicing principled negotiation

Clinical basic science role and academic medicine
- Clinical and service portfolio: understanding performance benchmarks
- Appointments and promotion
- Time management
- Conflict resolution
- Firing and hiring

Teaching and education
- Developing an education curriculum
- Giving feedback on professionalism

Leadership and policy
- Mentor and protégé relationships
- Leadership and team effectiveness
- Negotiation basics: practicing principled negotiation
- Negotiation advanced simulation: a retreat
- Translating research into policy and practice
- Policy series: legislative policy and health and fiscal policy
- Policy series: health and fiscal policy

independent study project (ISP), and faculty advisors for his ISP committee. The research mentor is responsible for the research and academic progress of the HCOE scholar. The scholar meets with his ISP chair on a regular basis (once per week during the period of high research activity) to discuss progress. The ISP chair advises the scholar on the selection of elective courses relevant to the research project, provides information on career opportunities within and outside academic medicine, and writes letters of recommendation and nominations for awards. The ISP chair is responsible for ensuring that the HCOE scholar receives training, experience, and assistance in writing scientific papers, giving oral presentations, and submitting research work for consideration at local, regional, and national meetings. Faculty advisors on the ISP committee evaluate the overall progress of the scholar, facilitate access to university and community resources, evaluate changes in research directions or problems, and review the final written summary of the research findings. Students are therefore learning first-hand to skillfully employ both the one-on-one and diffuse models of mentoring.

University of Pennsylvania School of Medicine–Penn Center of Excellence

As with all medical schools, the University of Pennsylvania School of Medicine has faculty research requirements that vary with the promotion track.[1,2] Some faculty interested in research have not had sufficient time allocated for research, the formal training needed for research, or the passion to succeed at the level of an NIH-funded physician scientist. Consequently, Penn has continually reinforced the necessity of matching one's activities with faculty track expectations and personal goals.

In order to define the expectations of each promotion track, the promotions committee chair meets with the faculty member and describes the value that Penn places on promotion criteria, such as the forms of scholarship (eg, original research, review articles, and case reports) appropriate to each track.[1,2] Subsequently, faculty interested in achieving the status of an independent investigator are given assistance with pilot funds, data analysis, funding for formal training in research methods, and appropriate mentoring by accomplished investigators.

With respect to its URM FDP, the Penn COE maintains a list of formal education opportunities for faculty and assists faculty in matching those opportunities with their career goals. Most basic science faculty members at Penn obtain doctoral training. For these faculty members, the selection of a laboratory and the selection of a particular field of basic science inquiry are concurrent. Many enroll in Penn's combined-degree program, which hosts one of the largest groups of URM combined-degree students in the nation. Faculty interested in clinical research, health services research, or translational research are apprised of other opportunities. Enrollment in the master's degree program in the Center for Clinical Epidemiology and Biostatistics is made available to any Penn minority faculty member at no cost. The Penn COE maintains a list of other course offerings on campus for persons with other career interests such as the Masters in Public Health Program and the Masters in Translational Research Program. Faculty members who wish to collaborate as investigators (referred to at Penn as clinician educators), rather than develop an independent investigative career on the model of the NIH-funded physician scientist, are offered selected introductory-level, non–degree-granting courses specifically designed for this level of investigator.

Several URM faculty members have expressed an interest in health disparities research, and some of those have been interested in community-based participatory research. Because of the COE's integral role in a vast array of health disparities centers across the school of medicine and other schools of the campus, the COE has been able to direct interested URM faculty to appropriate mentors and resources. Because the extensive time devoted to building trusting relationships in order to conduct community-based participatory research can be an impediment to promotion, especially when publications are expected early in one's career, the COE assists interested URM faculty by pointing out those community partners that have established relationships with Penn researchers and identifying senior faculty brokers in the relationship between the junior faculty and the community organizations.

Personal Positioning and Planning

Personal positioning and planning refers to a junior URM faculty member's process of examining, assessing, identifying, and implementing strategies conducive to successful advancement within the academic medicine environment. For many junior URM faculty members, personal goals may include undertaking more community-based or community-service work, developing expertise in another specific area of interest, and starting a family or spending more time with one.

The process of personal positioning and planning must be initiated by the junior URM faculty member himself; otherwise, the need may go unmet. The faculty member needs to take an active role in identifying those personal traits and goals that are compatible with those of the institution and in using the institution's tools to aggressively pursue and implement the joint plan. These actions will facilitate the creation of an environment beneficial for both the faculty member's career advancement and the institution's mission.

Mount Sinai School of Medicine Center for Multicultural and Community Affairs

The Mount Sinai CMCA program regards a faculty member's positioning and planning as fundamental. Embodied in an individual faculty development plan (IFDP; Table 3), positioning and planning are the subject of an ongoing dialogue initiated and promoted by the FDP staff. The IFDP is thus always a work in progress. It is not meant for submission to a division chief or department chair but belongs to the junior faculty member whose goals are articulated within it. However, the IFDP guide does include worksheet pages that can be copied, completed, and attached to a feed-forward meeting or appended to an annual report.

University of Medicine and Dentistry of New Jersey–New Jersey Medical School

When junior URM faculty members first join the New Jersey Medical HCOE, they are asked to complete a faculty development questionnaire. In addition to informing the HCOE about the members' goals, the questionnaire asks them to consider and clarify their values with respect to their personal and professional lives. For example, new faculty members are asked

Table 3. *Mount Sinai Center for Multicultural and Community Affairs Individual Faculty Development.*

Preparing yourself
- Assess your personal and professional time commitments, including current and desired percentages of effort invested in them.
- Identify what you value and broad goals that you have in mind for 3 to 5 years.
- Articulate the specific goals you have in teaching and learner development, scholarship, research and creative achievement, patient care and clinical leadership (for clinicians), university service, and university administration.
- Identify skills needed to successfully accomplish your goals.
- Identify your mentor coach for each area, an individual who will give you constructive feedback and/or advice.

Individual faculty development plan
- Specific goals
- Percentage of time committed to the area of your work
- Objectives
- Tasks and timelines to be addressed each quarter
- Anticipated outcomes
- Achievements by the end of the year
- Action items to put on your master list and calendar

Consulting with your mentor or coach
- Use this section to summarize the professional goals worksheet, including the current and desired allocation of your time, specifics about your academic appointment, and resulting achievements in teaching, research, and service (clinical, teaching, and administration, as defined).
- Set the agenda for a meeting with your mentor/coach. Be specific about goals, objectives, barriers, and desired feedback.

Standardized curriculum vitae format and instructions
- A sample curriculum vitae and instructions are included.

Meeting with your mentor or coach
- Agenda
- Individual faculty development plan forms (feed-forward)
- Annual review forms (feedback)
- Curriculum vitae (up to date)

to rate the importance of being role models and of balancing professional demands versus personal demands as well as the difficulty of declining a possible request to perform committee work and community service activities. They are also asked to describe their strengths in education, administration, research, written communication, mentoring, and service to others and the intensity of their motivation to conduct research, teach, and provide patient care or community service. Finally, the new members are asked to state their 1-, 3-, and 5-year goals and, on the basis of those goals, to identify areas in which they would seek additional training (HCOE Faculty Development Questionnaire, 2003)

University of Pennsylvania School of Medicine–Penn Center of Excellence

The Penn COE counsels junior URM faculty to do what they do best and pursue those career goals for which they hold the greatest passion. Within this framework, faculty are encouraged to be decisive in setting their goals and to take responsibility for accomplishing them.

As junior faculty members set their goals, they commonly reveal certain oversights and misunderstandings. For example, junior URM faculty members often miss opportunities to collaborate with one another on scholarly projects. Similarly, many junior faculty members in the 5 promotion tracks of the school of medicine are unclear about the expectations and requirements for promotion. Some of that misunderstanding derives from praise from department and division chiefs for clinical and administrative activities, although those activities might not be valued highly by the promotions committee (depending on the type of promotion track).

To correct these misunderstandings and oversights, the Penn COE has established an annual meeting of all new faculty with senior URM faculty and the Chair of the Committee of Appointment and Promotions. At this meeting, junior faculty members are informed of the key school and university-wide resources available to faculty (such as the Center for Clinical Epidemiology, the Leonard Davis Institute for Health Economics, and the Office of Faculty Affairs) and how to critique and participate in the development of their academic plan, which is a formal contract between each faculty member and the appropriate department chair that states faculty activities, the mentoring plan, and the resources available to support the faculty member.

MENTORING

Introduction

Importance of Mentoring

Medical FDP administrators are unanimous in their belief that mentoring is essential to the success of their programs.[2,4–10] Mentoring is necessary for a participant's

- Socialization into the field of medicine.
- Development of communication, teaching, and networking skills.
- Understanding of research methods.
- Relief from feelings of isolation.
- Skill in building collaborative networks and relating to students from diverse academic and cultural backgrounds (if the mentee aspires to a career as an educator).

In addition to its value for individual participants, mentoring is perceived as valuable to institutions.[11] Faculty mentors provide institutional leadership, embody the institution's intellectual capital, generate revenues, and hold its institutional memory. Moreover, by inducing faculty to remain, mentoring reduces the cost of recruiting and training new faculty.

On the basis of these perceptions, the American Association of Medical Colleges charged its membership in 1998 to begin concerted and systematic efforts to provide junior URM faculty with effective mentoring programs.

Although the perception of the importance of mentoring is widespread, studies concerning the impact of mentoring on careers in medicine, especially academic medicine, have so far been inconclusive.[9] In 2003, Jackson et al.[7] sought a deeper understanding of mentoring through interviews with 16 medical school junior faculty members. After exploring the subjects' lived experiences, the authors determined that having a mentor is critical to one's success in academic medicine. However, the generalizability of their findings is limited by the study's qualitative methodology and small sample size. In 2002, Ramanan et al.[12] explored mentoring relationships among residents in internal medicine. The study examined the relationship between mentoring and the residents' perceived preparation for medical careers. The authors designed and administered a survey mailed to all interns and residents enrolled in 5 independent internal medicine residency training programs affiliated with Harvard Medical School. Of the 329 respondents (65% response rate), 93% stated that it was important to have a mentor during residency training, half identified themselves as having a current or past mentor, and mentored residents were almost twice as likely to describe excellent career preparation. In 2006, Sambunjak et al.[9] conducted a systematic review of 39 studies that had evaluated the role of mentoring in academic medicine. The reviewers concluded that, given pervasive weaknesses in the study designs, the studies did not provide strong evidence for the perception that mentoring is an important part of academic medicine or permit a correlation of mentoring with success in an academic career.

To date, no studies have addressed the impact of mentoring specifically on the development of URM faculty.

Mentoring Differences Between Minority-Serving and Majority-White Institutions

Majority-white institutions and minority-serving institutions (MSIs) are the same with respect to most aspects of mentoring, but not all.

To the extent that the success of mentoring depends on similarities between the mentor and mentee with respect to race, ethnicity, and culture, junior URM faculty members may find suitable mentors more plentiful in MSIs than in majority-white institutions.[13] In MSIs, junior URM faculty members are surrounded by faculty and staff who are in synch with them culturally and fully understand the issues of racism and health justice. In majority-white institutions, junior URM faculty members may feel culturally and socially isolated, with relatively few senior faculty who can serve as role models and mentors, especially given the historic underrepresentation of minorities in academic medicine.[12]

On the other hand, mentors in majority-white institutions may be able to offer mentees more professional advantages than those in MSIs. Compared with MSI mentors, mentors in majority-white institutions may have a greater ability to add junior faculty to grant applications on which they are the principal investigators and to peer-reviewed papers on which they are the lead authors.

Mentoring Process

The mentoring process is complex and multifaceted.[2] Walker et al.[14] reviewed the numerous mentoring systems that have been proposed and determined that the common attributes are a knowledge differential between the participants, a teaching/learning process that involves reciprocity, and a career development relationship that lasts several years.

The variety and intensity of mentoring programs at US medical schools suggest the difficulty of successfully implementing mentoring programs that are formal. Mentoring often occurs informally, sometimes without a clear beginning or end, when 2 persons with a common interest simply find each other. Fagan and Walter[15] identify "definite mentoring" as a situation in which one person is especially influential and "diffuse mentoring" as a situation in which several people are influential.

Scope of Mentoring

Mentors must be knowledgeable in their area of expertise, wise about the institution in which they work, and capable of improving their mentees' research and teaching skills.

Equally important, mentors must provide general counseling and personal guidance and support. When asked to identify desirable characteristics of a role model, medical students, house officers, and physicians consistently include compassion as well as clinical skills and excellent teaching.[12] Junior URM faculty members require mentors who can provide them with advocacy, inspiration, socialization, and nurturing and do so with passion and authenticity.

Within this framework, mentors should encourage and help junior faculty to acquire

- Knowledge of the university history, organizational culture, and operations.
- Knowledge of the roles, responsibilities, promotion and advancement, performance review, and compensation of faculty.
- Personal communication, negotiation, and leadership styles that will influence the ability to be effective on committees as teachers and clinicians.
- Specific skills for academic roles in research, including formulating a research plan, writing grants and scientific manuscripts, identifying funding sources (NIH, foundations, and corporate sources), and competing for these awards.
- Specific skills for academic roles in teaching, including curriculum planning and design, student evaluations and feedback, and facilitation of small groups and presentations.

Mentors can also champion junior URM faculty by

- Writing letters of support for promotion.
- Offering guidance for applications for funding.
- Nominating junior faculty for local, regional, and national awards.
- Assisting in identifying resources.
- Recommending and nominating mentees for professional societies.
- Assisting in establishing professional networks locally, regionally, nationally, and internationally.
- Providing guidance in balancing work and personal life.

Finally, addressing institutional leaders, mentors can advocate the creation of a departmental and institutional climate that is supportive. URM faculty are often isolated within their departments, and this necessitates a serious effort to create strategies and structures to recruit and sustain them.

Mentor-Mentee Relationship as a Personal One

At its optimum, mentoring is a personal interaction between the mentor and mentee that promotes support, mutual respect, and compatibility.[2,7] Transcending simple role modeling, it provides the mentee with psychosocial and career support that takes account of the mentee's core values and priorities.

A mentoring relationship requires that the mentor and mentee choose each other. Communication between the two is facilitated if they share an interest in the same scholarly work or face similar career challenges. Some FDPs favor mentors and mentees who are of the same race, ethnicity, and/or gender, but available evidence suggests that mentoring can succeed whether they are the same or different.[13,16]

To build a relationship of trust, both the mentor and mentee must take responsibility for

- Meeting according to a prearranged schedule.
- Negotiating and clarifying the expectations of the relationship.
- Clearly articulating goals and objectives.
- Following through on commitments.
- Respecting gender, racial/ethnic, and other differences.
- Terminating the relationship when that is appropriate.

This interaction must also provide a safe space for honest discussion about personal issues that have the potential to disrupt the career of the mentee. One possible issue is resistance to mentoring. Mentoring includes socialization into the field of medicine, which is intended to provide a mentee with networking skills, fruitful peer relationships, and support for a successful career as a clinician, teacher, and/or investigator. Despite these benefits, a mentee may perceive socialization as a process of assimilation that lends itself to homogeneity and risks surrender of the mentee's personal identity.[17,18] Resolution of such a conflict can be instructive for both the mentee and the mentor.

Mentors are thus encouraged to be aware of the challenges faced by junior URM faculty and treat them with patience and due respect for differences. Such a process provides fertile academic soil not only to enrich the academic success of the mentee but also to challenge mentors to remain confident, comfortable, and proactive supporters of campus diversity and welcome the unique contributions of diverse faculty members.[16]

One-on-One Mentoring Versus Team Mentoring

The traditional view of mentoring is an interaction between 2 individuals, the master (mentor) and protégé (mentee), the latter developing in the image of the former.[15] The success of this replica development depends on the capabilities of the master, the example he or she sets, and the mutual trust in the relationship.

This approach to mentoring can sometimes be successfully used in academic medicine, but not always. Even if the necessary trust is assumed, the demands of the relationship may exceed the capabilities of a single mentor. In addition to performing the numerous and multifaceted tasks described previously, a mentor must meet the demands of the academic institution as a whole and the community in which the institution resides. Moreover, as time passes, both the needs of the junior faculty member and the educational culture and environment may change.[1]

Consequently, mentoring may require interaction between the mentee and more than 1 mentor, either concurrently or over time. Such a team approach is often the optimum way to fulfill the demands of the mentorship relationship. What it may lose in continuity, mutuality, and role modeling, it may gain in flexibility, speed of response, and resources.

Given a team approach, 1 responsibility of each mentor is to encourage and help junior faculty members identify other mentors who can support their career development, providing them with additional expertise, resources, and connections as necessary.

Institutional Examples

Michigan State University Center of Excellence for Cultural Diversity in Medical Education

The Michigan State University COE has established a mentoring process that uses multiple mentors. This model stemmed from the recognition that a senior faculty member, although having various academic capabilities, may not have sufficient experience and understanding of the cultural and social barriers facing a URM faculty member to fully support the latter's development.

A COE mentoring team thus consists of a senior, departmental mentor identified by the department chair and a second mentor who provides a complementary cultural or organizational perspective and may come from a different discipline. The departmental mentor serves the role of research collaborator, reviewer of scholarly work, and guide to understanding the complexity of the academic promotion process. The complementary mentor provides linkages with a wider network of university professionals, possibly including opportunities for multidisciplinary collaboration. Acting as a team, the mentors meet with the mentee on a monthly basis to review goals and progress, modeling a collaborative process.

Mount Sinai School of Medicine Center for Multicultural and Community Affairs

The Mount Sinai CMCA provides both mentoring and coaching. Mentoring is similar to mentoring in other institutions and tends to occur at regular intervals. Coaching is less formal, providing continuous, cross-cutting, and coordinated support for other components of the FDP. Coaching thus provides the glue between other FDP components.

Coaches may or may not also be mentors but in any case possess deep experience in the institutional culture and environment. As such, they help junior URM faculty members

- Take a complete look at their current state, including their assumptions and perceptions about their work, themselves, and others.
- Set relevant and realistic goals for themselves that are based on their own nature, needs, and goals.
- Take relevant and realistic actions toward reaching their goals.
- Learn by continuing to reflect on their actions and sharing feedback with others along the way.
- Hold themselves accountable for doing what they said they would do by when they said they would do it.

Conducted on both a one-on-one basis and a group basis, coaching is a time-consuming process. For efficiency, it is best performed at the same time that a junior faculty member engages in personal positioning and planning.

One example of coaching involves mentors and sponsors who have differing expectations and timeframes regarding the faculty development process. To coordinate their efforts, the CMCA has arranged for the FDP staff to meet with junior faculty members on an individual basis and for both the staff and participants to meet as a group at monthly 1-hour breakfast meetings. Involving interdisciplinary collaboration, the latter meetings evaluate both progress on current research projects and overall personal and professional development as it pertains to the IFDP.

University of Arkansas College of Medicine

The University of Arkansas College of Medicine has developed a mentoring program for junior URM faculty called the Peer-Onsite-Distance (POD) program. The POD program seeks to both retain developing URM faculty members and advance their careers. It targets each faculty member's career goals, resources, and content and interpersonal skills needed for a successful academic career, while providing the member with a multilevel network of mentors. The network includes peer and faculty mentors who provide site-specific career guidance and others off-site who can provide accurate information, cautions, predictions, and announcements of future resources or potential restrictions in particular fields of academic medicine. The POD program thus wraps each member in a protective, multilayered cushion of interpersonal support.[19]

University of California at San Diego Hispanic Center of Excellence

Collaborating with the National Center for Leadership in Academic Medicine, the University of California at San Diego HCOE has designed an FDP whose cornerstone is a formalized, proactive, instrumental mentoring process.[6] Mentees participate in workshops and counseling sessions, the total duration of which exceeds 150 hours per year. During this time, senior colleagues collaborate with junior faculty on research or teaching projects, critique their scholarly work, and nominate them for career-enhancing awards and other specified tasks.

The HCOE also pursues faculty development by inviting Hispanic medical students to participate in the HCOE Scholars Program, during which the students engage in structured research under the guidance of senior investigators.[6] The projects fall into 3 major categories: basic science research, clinical research, and applied research in community settings. HCOE scholars may also engage in health promotion and the evaluation of programs designed to improve health literacy and health indicators in underserved communities. In addition, scholars are encouraged to establish collaborations with community partners that include converting their activities into research projects that benefit both the community and the university.

The HCOE Scholars Program begins in the second year of medical school with the scholar's selection of a research mentor, ISP, and faculty advisors for his ISP committee.[6] The research mentor is responsible for the research and academic progress of the HCOE scholar. The scholar meets with the ISP chair on a regular basis (once per week during the period of high research activity) to discuss progress. The ISP chair advises on the selection of elective courses relevant to the research project, provides information on career opportunities within and outside academic medicine, and writes letters of recommendation and nominations for awards. The ISP chair is also responsible for ensuring that the HCOE scholar receives training, experience, and assistance in writing scientific papers, giving oral presentations, and submitting research work for consideration at local, regional, and national meetings. Faculty advisors on the ISP committee evaluate the overall progress of the scholar, facilitate access to university and community resources, evaluate changes in research directions or problems, and review the final written summary of the research findings.

HCOE scholars thus learn first-hand how the one-on-one and diffuse models of mentoring can be skillfully employed.[6] As such, the HCOE Scholars Program is a fertile field for developing the next generation of Hispanic faculty.

University of Pennsylvania School of Medicine–Penn Center of Excellence

In 1994, the Penn COE undertook a study to learn the extent of mentoring programs in the departments and divisions of the school of medicine and to compare URM and non-URM faculty members with respect to their mentoring experiences.[1,2] As a result of this information, the COE established a mentoring selection and advising program for URM faculty. Subsequently, the URM mentoring program itself became the basis for greatly bolstering faculty mentoring programs overall.

The Penn COE URM mentoring program is based on a list of potential faculty mentors delineated according to the type of research expertise: basic science, clinical epidemiology, health services delivery, and community-based participatory research.[1,2] The COE shares this list with all entering URM faculty and holds one-on-one meetings with faculty about their choice of mentor or mentors. In addition, the COE has established a short list of senior minority basic scientists to whom junior faculty (fellows and postdoctorate trainees) are referred for advice about the various laboratories across the campus.

The Penn COE provides each URM faculty member with a checklist of steps to take on an annual basis, such as meeting with department chairs and division chiefs, maintaining a record of teaching assignments and the trainees in those assignments, and lectures by invitation.[1,2] Once a year, the COE reminds each faculty member to update the annual checklist and complete a survey that allows junior faculty members to critique the mentoring that they have been receiving. COE leaders then assist the faculty members in rectifying problems with the mentoring process. In addition, COE staff members with expertise in research methods are available for consultation in research design and assistance with establishing a research database and data management system, data analysis, and manuscript preparation.

SELECTING PARTICIPANTS

Trainees

Albert Einstein College of Medicine

At the Albert Einstein College of Medicine, faculty development fellowships have been offered to new candidates under the organizational umbrella of the Institute for Community and Collaborative Health (ICCH). ICCH programs include the HCOE funded by the Health Resources and Services Administration (HRSA), the NIH-funded Bronx Center To Reduce and Eliminate Ethnic and Racial Health Disparities (Bronx CREED), and the Josiah Macy Jr. Foundation–funded NECMFD. The fellowships share Einstein's NIH-funded K-30 master's degree program, the CRTP, and a common application to both programs.

Fellows must be selected by both the CRTP Admissions Committee and the ICCH Selection Committee, which includes hospital-based, university-based, and community-based URM faculty. The selection criterion are as follows:

- A clinical doctoral degree (eg, MD, DDS, DO, PharmD, PsyD, or PhD).
- Strong interest in and aptitude for clinical research.
- Interest in improving healthcare disparities in minority populations.
- Appointment as a faculty member at an Albert Einstein College of Medicine–affiliated institution for the 2 years of the fellowship.
- A firm commitment from the fellow's department (eg, chair or division chief) that all necessary resources will be available, including office space, a computer, and protected time (ie, a reduction in clinical responsibilities).

Faculty development fellows receive full-tuition support, and their departments receive a fixed dollar stipend for salary and fringe benefits (approximately 30%–40% of a full time equivalent) for 2 years in order to complete the master's program. All fellows are required to have 50% effort protected time from their departments for the fellowship and CRTP. Fellows also attend monthly seminars on health disparities and faculty-wide workshops on academic skills, case-based teaching, and cultural competency. Thus far, the ICCH has supported a total of 12 fellows: 5 Latino HCOE fellows, 6 Bronx CREED fellows (2 African Americans, 1 Latina, 1 Egyptian American, 1 Iranian American, and 1 Caucasian), and 1 Macy fellow (who is Haitian-American).

Morehouse School of Medicine

To date, the Morehouse executive and modular FDP programs have been completed by 197 participants. Most have been junior faculty (instructors and assistant professors), teaching physicians, and non-MDs (nurses, dieticians, psychologists, and social workers) from primary care departments and the extended medical community. The program has also been completed by most new Morehouse faculty in family medicine and pediatrics as well as some chief residents and many community preceptors who teach Morehouse residents and students. A few basic scientists have participated, but HRSA has required the program to focus primarily on clinical faculty.

To attract applicants to its FDP, Morehouse appeals to those interested in issues unique to African Americans and other minority populations. Most applicants already work with underserved and disadvantaged populations. Although not all participants are people of color, all are dedicated to eliminating health disparities, providing culturally competent care, and being effective teachers and mentors to URM faculty and students at the Morehouse School of Medicine or at their home institutions.

In addition to healthcare professionals, the Morehouse URM FDP is administered to fourth-year medical students enrolled in the Academic Medicine elective. The purpose of this elective is to increase the pipeline of URM medical students who choose careers in academic medicine.

Mount Sinai School of Medicine Center for Multicultural and Community Affairs

Sources for Mount Sinai FDP candidates are the faculty pipeline, CMCA Faculty Advisory Council, and Internal Oversight Committee, as well as affiliates and partners of the Mount Sinai School of Medicine. Each candidate is selected on the basis of a strong potential for becoming a research scientist and a commitment to health disparities research, as demonstrated during an application process that includes the candidate's

- Curriculum vitae.
- Personal statement.
- Research summary that describes research career goals and interests.
- Areas of clinical expertise and willingness to contribute to health disparities research.
- Proposed research plan for the specific research topic and timeline.
- Letter of support from a department chair or division chief and proposed research mentor.

University of Minnesota Medical School

Because the number of American Indian faculty is small and research with and in American Indian communities requires extensive skills, URM faculty development is less of a reality at the University of Minnesota Medical School than elsewhere. To initiate an FDP, the University of Minnesota has brought together interested Native American residents as fellows in training. Several times annually, it assembles a critical mass of like-minded participants for education and experience in issues affecting the success of junior faculty, including time management, work with Native American communities, project development, cross-cultural medicine, and learning styles.

Teachers

Morehouse School of Medicine

Morehouse teachers who participate in its executive and modular FDP programs are a culturally diverse group of men and women who are all excellent teachers as well as experts in the subjects that they teach. Most FDP teachers are themselves graduates of the Morehouse program. Representing diverse departments–Family Medicine, Pediatrics, Psychiatry, Medical Education, and Community Health and Preventive Medicine–they also have a broad range of relevant and timely topical interests, including but not limited to cultural competence, childhood obesity, maternal and child health, depression, health literacy, and fetal alcohol syndrome.

Mount Sinai School of Medicine Center for Multicultural and Community Affairs

To recruit teachers for the FDP, the Mount Sinai CMCA consults with senior research and career development experts representing groups underrepresented in medicine, including not only racial and ethnic groups but also women. Located both inside and outside Mount Sinai, the consultants are almost all senior-level female scientists and clinicians.

DEVELOPING SENIOR URM FACULTY

Faculty Leadership

In addition to being few in number, URM academic faculty members may lack leadership skills. Promotions are often based not on leadership experience and training but on clinical accomplishments, exemplary teaching skills, and/or an ability to attract research dollars to the institution. An advancing URM faculty member may thus take over a department or division without the ability to manage it successfully, including the ability to attract more URM faculty members to the institution.[20]

Consequently, FDP administrators must answer 2 questions: how to retain mid-level URM faculty members and thus increase the chances of advancing them to senior positions and how to provide such faculty with the leadership skills that they need. Unfortunately, case-based literature addressing these questions is nonexistent. A review of the literature shows no structured or organized approach to the development of mid-career URM faculty as leaders.

Such approaches can, however, be found in non-URM literature. One example is a program for women, Executive Leadership in Academic Medicine (ELAM), which provides women with extensive leadership training and coaching as well as networking and mentoring opportunities.[21] Like URM programs, ELAM fills a specific need for diversification, that is, the diversification of medical division or section chiefs, 80% of whom are male.[22] As stated on the ELAM Web site, the rationale for ELAM is similar to that cited for URM development programs: "Placing more women in positions of senior leadership at AHCs [academic health centers] will provide important new perspectives for decision making and help speed the curricular, organizational, and policy changes needed to ensure a more effective, representative, and responsive healthcare system." The ELAM curriculum develops both the professional and personal skills needed to effectively lead and manage today's academic institutions.

In addition to programs for the healthcare professions, the relevant literature includes programs for business. Recently, business approaches to management development have undergone a reevaluation.[20] Rather than relying on a Darwinist selection process, which results in a choice of leaders from among those who have survived, businesses have begun implementing a new approach that involves the development of those who hold promise as future managers. This approach reflects a belief that the acquisition of leadership skills advances not only a person's career but also the mission of the institution. Appealing to both current and future leaders of the institution, it promotes change at both the individual and structural levels.

Northeast Consortium for Minority Faculty Development

The NECMFD regularly hosts formal programs to target the development of specific skills for junior and senior URM faculty. Among these is a 2-day minority senior faculty leadership institute for associate professors and professors to prepare them for such roles as division chief, chair of a department, and dean. Topics have included the following:

- Role model panel discussion: senior leadership discussion of personal lessons learned.
- Assessing the institutional environment.
- Strategic planning.
- Delegating duties.
- Basic financing and accounting methods.
- Hiring, managing, and giving feedback to staff.
- Recruitment and retention strategies for minority faculty.
- Evaluation, self and external.

Reducing Faculty Burden

In addition to responsibilities common to all medical faculty, mid-career URM faculty are often pressed to tutor and mentor junior URM faculty and serve on committees in order to represent minority views, although none of these extracurricular activities provides the mid-career faculty with career-advancing recognition.[23,24]

The extracurricular burden on individual URM faculty members decreases only when the URM faculty reaches a critical mass. Up to this point, the burden of expanding the faculty falls on a few mid-career members; after the critical mass has been achieved, the burden of further expansion, as well as other extracurricular activities, is divided among many URM mid-career members.

PROVIDING AN ENVIRONMENT CONDUCIVE TO SUCCESS

Teachers Who Are Nurturing

Morehouse School of Medicine

The Morehouse School of Medicine provides a nurturing, learner-centered environment. In a recent survey, past participants in the Morehouse FDP described this environment as:[25]

- "Not only were teachers always available, but they also showed individual interest, gave encouragement, and were accepting, of everyone."
- "One-on-one instruction was given when needed."
- "The learning process included demonstrating what was learned."

In addition, Morehouse provides faculty and graduates who are role models of success, prompting participants to feel "I can do that!" Participants thus experience a safe, inspirational environment in which they can build confidence and competence in their academic skills.

Cultural Match

Morehouse School of Medicine

The Morehouse URM FDP faculty consists of racially, ethnically, culturally, and gender-diverse clinicians' who serve as positive role models and attract participants who are similarly diverse. Participants therefore meet instructors and peers who have faced similar challenges and demonstrated resilience.

This cultural match permits an easy exchange of information, knowledge, and experiences that enriches the entire program. The URM FDP curriculum includes a session entitled "Cross Cultural Issues in Teaching and Learning," and weaves activities throughout the curriculum that build the students' skills in cultural competence.

Facilitated Learning

Albert Einstein College of Medicine

Under the ICCH, the HCOE and Bronx CREED joined the Division of Education in cosponsoring annual full-day retreats for women and minority faculty to acquire negotiation and teaching portfolio skills. The HCOE and Bronx CREED also sponsored a series of half-day workshops to develop individual pedagogical skills of URM and other faculty. Entitled "Using Cases to Teach Cross-Cultural Patient Care: Train-the-Trainer Workshop," the series consisted of 2 workshops, the second building upon the first.

Following the train-the-trainer workshops, the participating faculty prepared a monograph, *Teaching Cases Exploring Cross-Cultural Care*, that was distributed to faculty and health disparities centers and COEs around the country.[26] The workshops and the Einstein-wide Faculty Task Force on Cross-Cultural Training also provided the basis for a faculty speakers bureau.

Morehouse School of Medicine

The Morehouse FDP program empowers participants with skills that increase their productivity in areas of greatest importance for academic medicine, such as writing papers for publication and preparing applications for grants. Each workshop is a facilitated learning session that involves active coaching, not simply a lecture, and practice of a needed skill, not simply hearing about it. Participants thus complete each module with a real-life product or activity, such as a letter to the editor, a concept paper for a grant, a health services research poster, or a large-group presentation with PowerPoint.

Buying Out Protected Time

Mount Sinai School of Medicine Center for Multicultural and Community Affairs

The Mount Sinai CMCA offers a distinctive approach to URM faculty development through its Faculty Scholars Program (FSP). The FSP is a professional development program that increases the capacity of URM faculty by buying out protected time for their development. Taking the form of a salary stipend, the buyout also serves as an incentive to department chairs to support both the protection of development time and the recruitment of additional URM faculty.

Comprehensive in scope, the FSP enables junior faculty members to immerse themselves in research and technical activities that prepare them for senior academic positions, including research, teaching, education, clinical practice, and professional academic leadership. At the same time, the junior faculty participate in formal and individualized just-in-time education career development activities with CMCA FDP staff, CMCA FDP consultants, and their respective research and career development mentors.

University of Medicine and Dentistry of New Jersey–New Jersey Medical School

The New Jersey Medical School HCOE allocates funds for trainees' salary support, which provides protected time for additional training and research. An integral component of the program, this release time enables a junior faculty member to fulfill the goals stated in the member's faculty development plan.

University of Pennsylvania School of Medicine–Penn Center of Excellence

The Penn COE offers partial salary support to selected URM faculty scholars who aspire to become investigators or clinical educators.

Exposure to National Leaders

Albert Einstein College of Medicine

Sponsored by the HCOE and Bronx CREED, train-the-trainer workshops for URM faculty have been led by model teachers of national renown, including Dr. Debbie Salas-Lopez, Division Chief of General Internal Medicine at Lehigh Valley Hospital and Associate Professor at Penn State, and Dr. Joseph Betancourt, Assistant Professor of Medicine at Harvard and Director of the Health Disparities Solution Center at Massachusetts General Hospital.

Morehouse School of Medicine

The Morehouse FDP includes the frequent exposure of participants to national leaders in URM education. Examples include the following:

- Dr. David Satcher, 16th Surgeon General, former Director of the Centers for Disease Control and Prevention, and current Assistant Secretary of the Department of Health and Human Services, meets annually with each group of participants in an informal luncheon discussion.
- Dr. Edith Irby Jones, former President of the National Medical Association and the first African American to attend University of Arkansas Medical School, delivered an inspirational talk at the 2003 end-of-year celebration.
- Eve Higginbotham, Dean of the Morehouse School of Medicine and Senior Vice President for Academic Affairs, attended the Morehouse "Meet

the Leader" luncheon and spoke at the annual 2007 graduation.

- Morehouse School of Medicine President John Maupin met informally with the 2007–2008 group of participants and was the keynote speaker for the 2008 FDP graduation.

MANAGING THE PROGRAM

Institutional Support

To be successful, a URM FDP requires vigorous, broad-based institutional support for the goals of diversity. This support must be

- Continuous along the faculty member's movement from a junior rank to a senior rank.
- Cross-cutting with respect to all realms of the member's responsibility, including teaching, education, personal lifestyle, and community and institutional service.
- Coordinated with respect to time, attention, and resources (eg, salary support to provide protected time for career development).

Northeast Consortium for Minority Faculty Development

To attract and develop URM physicians and nonphysician scientists to FDP programs, the NECMFD members make a 3-year investment in each junior faculty member, including flexible local use of funds and specific funding for career development and pilot research activities.

Mount Sinai School of Medicine Center for Multicultural and Community Affairs

The Mount Sinai CMCA has found that the success of a URM FDP requires the values and priorities of diversity and those of the institution as a whole to be aligned. Despite turnovers in institutional leadership, the CMCA has been fortunate in having the consistent support of the leaders for over 10 years. For example, since 2004, a diversity statement prepared by the CMCA has been included in handbooks for medical students, postgraduate students, and faculty. The dean of the school of medicine has also regularly and consistently endorsed diversity as an institutional priority and the CMCA's work as important. This support has been a major facilitator in gaining CMCA the cooperation of department chairs and other academic leaders.

University of Pennsylvania School of Medicine–Penn Center of Excellence

The success of the Penn COE would not have been possible without the commitment of the top administrative leadership of the medical center and without the interest of diverse institutes and centers of the university in minority health and career development.[1,2] This support has been shown in both funding and policies and structures that create a welcome environment for URM faculty. One example was the establishment in 2005 of the Office of the Senior Vice Dean for Diversity and Community Outreach.

University of Medicine and Dentistry of New Jersey–New Jersey Medical School

The success of the New Jersey Medical HCOE has depended on the institution's commitment and willingness to support a culture that promotes diversity. The development of each trainee is overseen by the HCOE director. Because the director has key decision-making capacity at the institution, the oversight provides trainees with a facilitating environment. The director listens to their needs, provides them with feedback, gives them the tools to move forward, champions their issues, causes, and successes, and reinforces advice from chairs or division directors.

In providing oversight, the director has received assistance from senior faculty members in the form of skilled and experienced feedback and help in training and mentoring each trainee. At the outset, the HCOE FDP was supposed to include a senior faculty advisory committee that would meet with the director on a regular basis and maintain a relationship with each trainee's department chair. However, given the seniority of the committee members, regular meetings proved difficult logistically. Instead, senior faculty members have assisted the director on an individual basis. With their invaluable assistance, the FDP has succeeded in achieving many of its goals.

Strategic Positioning

Mount Sinai School of Medicine Center for Multicultural and Community Affairs

The success of the Mount Sinai CMCA has depended importantly on its strategic positioning within the institution's organizational structure. Prior to HRSA funding as a COE, the CMCA operated as an office reporting to the Dean of Student Affairs. After HRSA funds became available, medical school leaders agreed that, in order for the CMCA to have an optimal impact on institutional diversity, it would need to engage in cross-institutional collaboration and report directly to the medical school dean. In addition to this direct reporting relationship, the CMCA is now centrally located in the school of medicine.

Independent Governance

Mount Sinai School of Medicine Center for Multicultural and Community Affairs

Although accountable to the medical school dean in meeting designated benchmarks, the Mount Sinai CMCA has relative autonomy in establishing diversity program priorities and policies.

The CMCA shares these responsibilities with an advisory board composed of internal and external partners.

The CMCA also receives assistance and advice from an internal oversight committee of deans and chairs. This committee helps the CMCA work with existing departments to enhance the institution's ability to achieve its core missions. The committee's priority is to advise and assist the CMCA with the implementation of diversity initiatives involving both faculty and house staff.

Pipeline Improvements

The pipeline for URM faculty begins in high school and extends through undergraduate and graduate studies into the professional period. So far, each segment of the pipeline has been deficient, "leaking" potential or actual members of the URM faculty. Although authorities differ on which leaks are most important to plug, they agree that all segments of the pipeline require strengthening.

In 2003, Hallock[27] examined the perceptions of academic life among 17 URM students participating in either the Alliances for Graduate Education and the Professoriate or the Ronald E. McNair Post Baccalaureate Achievement Program. The author found that many of the students had formed negative perceptions of academia and developed coping mechanisms accordingly. The primary coping mechanism was "separating the personal," which involved efforts to reconcile experiences of individual merit and achievement with those of discrimination or marginalization. Striving to navigate and negotiate the route to an academic career, they felt at risk openly expressing their personal opinions of academic norms. The study thus confirmed that students form their impressions of academic life at an early stage and that these impressions have an important and sometimes negative impact on their pursuit of academic careers.

Mount Sinai School of Medicine Center for Multicultural and Community Affairs

Adopting a growing-our-own approach, the Mount Sinai CMCA cultivates the growth of students interested in medical academic careers, starting as early as junior high school and extending through undergraduate, medical school, resident, and postgraduate settings. The CMCA has found that this long-term investment provides a significant return in later years.

University of Pennsylvania School of Medicine–Penn Center of Excellence

Because of the limited number of URM faculty nationwide and because a redistribution of faculty from another institution to the University of Pennsylvania would not best serve the nation's interest, the Penn COE has embraced a growing-our-own philosophy. The COE leadership espouses the view that preparation for a faculty career should begin during the undergraduate years, if not earlier, and continue into medical school, involving MD and combined-degree (MD-PhD) students. Because most of the minority faculty at the University of Pennsylvania (and elsewhere) hold MD degrees rather than PhD or combined degrees, a large pool of potential faculty is overlooked by research-intensive medical centers that concentrate solely on combined-degree students.

Since 1991, the University of Pennsylvania has offered a summer enrichment program, which is now 10 weeks in duration. The principal goals of the program are to generate interest in research among URM undergraduate students who have had limited exposure to research and to prepare students for academic careers in medicine. URM students, predominantly but not exclusively from historically black colleges and universities, are taught research skills through a combination of didactic sessions and participation in research projects under the guidance of faculty mentors. Other aspects of the program include basic science lectures, clinical preceptorship, and preparation for the Medical College Admission Test. At the end of the program, each student is expected to give a 30- to 45-minute presentation pertaining to his work.

Affirmative Action Alternatives

States such as California and Michigan that have explicitly banned affirmative action must find acceptable, effective, and legal alternatives for promoting diversity.[28] This will require rethinking the criteria and processes that health professions schools use for student admissions and faculty recruitment. In place of racial and ethnic categories, the schools may need to use cross-cutting criteria, such as success in overcoming educational disadvantage, commitment to and experience in underserved communities, and demonstrated interest and expertise in studying and reducing health disparities.

SUSTAINING SUPPORT

Introduction

The long-term success of URM faculty development depends on sustained public, foundation, institutional, and individual commitments to establishing an academic health professional workforce as diverse as the clinical populations that it serves.

Political and institutional efforts and individual and collective action are needed at all levels to ensure that health professions schools and training programs are filled with talented faculty of color. The evidence that existing FDP programs have been effective at every level of the health career pipeline must be disseminated in the mainstream health professions literature. To date, it has been ghettoized in Institute of Medicine, federal, and foundation reports, where it can easily be ignored or dismissed by those who oppose diversity programs.[29–31]

In this period of considerable financial distress and changing paradigms for academic careers, health professions schools must honestly examine not only their policies for recruitment, support, and promotion but also their organizational cultures and attitudes regarding faculty of color. If health professions schools, stressed by economic constraints, the impact of regress and fall into old habits and exclusive ways, federal and state legislation and funding will diminish.

Federal Support

Recently, Senator Edward Kennedy (D-MA) and Representative Jesse Jackson, Jr. (D-IL), with bipartisan cosponsors, introduced the Minority Health Improvement and Health Disparity Elimination Act through identical companion bills (S. 1576 and H.R. 3333, respectively) designed to improve the health and healthcare of racial and ethnic minority groups. Specifically, the bills address health disparities through education and training, healthcare quality and access, research, data collection, leadership, and collaboration. The bills focus on the following areas:

- Improving cultural competency among healthcare providers.
- Increasing diversity in the healthcare workforce.
- Enhancing healthcare quality and access.
- Research on health disparities and addressing genetic variations and health.
- Data collection, analysis, and quality.
- Leadership, collaboration, and a national action plan.

Among the significant components of the bill is the reauthorization of both COEs and Health Careers Opportunity Programs (HCOPs). Both bills have been referred to their respective committees and await hearings and further action. Reauthorization, however, does not guarantee adequate appropriations, so Congress needs to hear continually from health professionals, health professions schools, and the public about the importance of these programs to achieve a diverse workforce and eliminate health disparities. These efforts are supported by large coalitions, such as the Health Professions and Nursing Education Coalition, which is coordinated by the American Association of Medical Colleges, and smaller groups, such as the National Council for Diversity in Health Professions, that lobby in Washington, DC for the COE and HCOP programs.

Strengthening the ties between federal programs, especially between the HRSA and NIH, should focus not only on students but also on faculty. For example, NIH training or loan repayment awards could give priority to institutions and individuals affiliated with COEs and/or HCOPs, just as the NIH Endowment Program requires institutional designation as a COE to determine which schools are eligible to apply. As currently structured, the NIH Endowment Program strictly limits eligible institutions by considering university systems as a whole rather than specific health professions schools within them and targets endowment income toward diversity programs rather than faculty-oriented academic infrastructure, such as endowed chairs, centers, divisions, and departments with missions to eliminate health disparities.

State Support

State programs that are comparable to HCOPs, such as New York's Science and Technology Entry Program and College Science and Technology Entry Program, deserve expanded support and, where possible, integration with HCOP-funded and COE-funded programs. State-wide collaborations, such as the Associated Medical Schools (AMS) of New York's postbaccalaureate program hosted at the University of Buffalo School of Medicine, strengthen the diversity of incoming medical school classes. This year-long, 20-seat, postbaccalaureate program receives students nominated by participating New York medical schools. If they maintain an adequate academic record (ie, a B average) and improve their Medical College Admission Test scores, they are automatically admitted to their nominating medical school. The AMS program receives state funding as well as COE and HCOP funding through the Albert Einstein College of Medicine. AMS also sponsors HCOP-like enrichment programs on several medical school campuses with its Science and Technology Entry Program funding.

The NY State Department of Health also sponsors the Empire Clinical Research Program to encourage recent residency graduates to begin academic careers in clinical research. These research fellowships are based at hospitals and academic health centers around the state and offer stipend support under the guidance of an established mentor. Such faculty pipeline programs have been coupled with NIH-funded K-30 training programs to equip young investigators with the knowledge and skills to pursue clinical research careers. The NY State Department of Health AIDS Institute also offers the Nicholas Rango HIV Scholars Program, a fellowship for primary care clinicians interested in strengthening their clinical knowledge and leadership skills in caring for HIV-positive individuals. This, too, can serve as an entry to an academic career for clinicians of color.

Private Support

The Robert Wood Johnson Foundation and W. K. Kellogg Foundation have supported several different programs for diversity in the health professions, including health career pipeline programs, health disparity and minority faculty development fellowships, and special reports, such as the Sullivan Commission. Both foundations have also funded the Health Professions Partnership Initiative administered by the AAMC as well as grant awards to 16 partnerships.

Similarly, the California Endowment has expanded its grant-making goals to include reductions in health disparities and the development of a culturally competent workforce and has sponsored reports and programs addressing health professions workforce diversity in California. The AAMC also continues annually to offer the ELAM program for both women and minority faculty.

Institutional Examples

Northeast Consortium for Minority Faculty Development

As noted earlier, the NECMFD is composed of 4 COEs previously funded by the HRSA: the Albert Einstein College of Medicine, Mount Sinai School of Medicine, University of Dentistry and Medicine of New Jersey–New Jersey Medical College, and University of Pennsylvania School of Medicine. The NECMFD is now funded by the Josiah Macy Jr. Foundation, which has a history of supporting health career pipeline programs. With foundation support, each member of the NECMFD continues to provide faculty development fellowships and pilot grants to its own faculty and conduct annual workshops for junior and senior minority faculty.

Minority-Serving Institutions

Often lacking high levels of endowment monies, MSIs are in special need of support from the HRSA, Office of Minority Health, state and local governments seeking to increase URM faculty to match the diversity of their populations, and/or other public or private sources.

MSIs need such funds for both the programs themselves and faculty time to allow participation. Compared with their white counterparts, URM clinical faculty often spend greater amounts of their time in clinical practice or service delivery. Although this is consistent with the motives of many URMs in entering medicine, it is not consistent with success in academic medicine. Time spent in clinical practice is time not spent in learning how to pursue or publish research or write grant proposals. With salary support to free them from clinical responsibilities during development, URM faculty can learn how to compete successfully for grant funds later and thus eventually provide for their own support.

The appeal of MSIs for such funds is based on the training that it provides in common with other schools, such as courses in writing and health services research, and the focus of MSI programs on eliminating health disparities, which resonates well with the interests of most URM junior faculty. MSIs thus bring extra, intangible qualities to the faculty development experience.

SUMMARY

This article describes the ingredients of successful FDPs for minorities underrepresented in academic medicine. Although stung by recent cuts in federal funding, minority FDPs now stand as models for medical schools that are eager to join the 140-year-old quest for diversity in academic medicine.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Johnson JC, Jayadevappa R, Taylor L, et al. Extending the pipeline for minority physicians: a comprehensive program for minority faculty development. *Acad Med* 1998; 73: 237–244.
2. Johnson JC, Williams B, Jayadevappa R. Mentoring program for minority faculty at the University of Pennsylvania School of Medicine. *Acad Med* 1999; 74: 376–379.
3. National Institutes of Health. K30 Clinical Research Curriculum Award (CRCA). Available at: http://grants2.nih.gov/training/k30.htm. Accessed September 2008.
4. Accreditation Council for Graduate Medical Education. *Graduate Medical Education Director, 2008–2009.* Chicago, IL: American Medical Association: 2008.
5. Cohen JJ. Time to shatter the glass ceiling for minority faculty. *JAMA* 1998; 280: 821–822.
6. Daley S, Wingard DL, Reznik V. Improving the retention of underrepresented minority faculty in academic medicine. *J Natl Med Assoc* 2006; 98: 1435–1440.
7. Jackson FA, Palepu A, Szalacha L, et al. Having the right chemistry: a qualitative study of mentoring in academic medicine. *Acad Med* 2003; 78: 328–334.
8. Kosoko-Lasaki O, Sonnino RE, Voytko ML. Mentoring for women and underrepresented minority faculty and students: experience at two institutions of higher education. *J Natl Med Assoc* 2006; 98: 1449–1459.
9. Sambunjak D, Straus SE, Marusic A. Mentoring in academic medicine. *JAMA* 2006; 296: 1103–1115.
10. Wilkerson W, Irby D. Strategies for improving teaching practices: a comprehensive approach to faculty development. *Acad Med* 1999; 73: 387–396.
11. Wingard D, Garman K, Reznik V. Facilitating faculty success: outcomes and cost benefit of the UCSD National Center of Leadership in Academic Medicine. *Acad Med* 2004; 79(suppl 10): S9–S11.
12. Ramanan RA, Phillips RS, Davis RB, et al. Mentoring in medicine: keys to satisfaction. *Am J Med* 2002; 112: 336–341.
13. Wesson DE, Kine TE Jr, Todd RF, et al. Achieving diversity in academic internal medicine: recommendations for leaders. *Am J Med* 2006; 119: 76–81.
14. Walker W, Kelly P, Hume R. Mentoring for the new millennium. *Med Educ* 2002; 7: 1–7.
15. Fagan M, Walter G. Mentoring among teachers. *J Educ Res* 1982; 76: 113–118.
16. Stravaraka C. Mentoring minority faculty: how crossing racial lines helps minorities succeed. *Hisp Outlook Higher Educ* 2005; 16: 24.
17. Bussey-Jones J, Bernstein L, Higgins S, et al. Repaving the road to academic success: the IMeRGE approach to peer mentoring. *Acad Med* 2006; 81: 674–679.
18. Girves J, Zepeda Y, Gwathemy J. Mentoring in a post-affirmative action world. *J Soc Sci* 2005; 61: 449–479.
19. Lewellen-William C, Johnson VA, Deloney LA, et al. The POD: a new model for mentoring underrepresented faculty. *Acad Med* 2006; 81: 275–279.
20. Hill L. Leadership development: a strategic imperative for higher education. Available at: http://net.educause.edu/ir/library/pdf/FFP0506S.pdf. Accessed September 2008.
21. Executive Leadership in Academic Medicine. ELAM program. Available at: http://www.drexelmed.edu/elam/aboutelam/index.html. Accessed September 2008.
22. Association of American Medical Colleges. Women U.S. academic medicine statistics, 2006–2007. Available at: http://www.aamc.org/members/wim/statistics/stats07/start.htm. Accessed September 2008.
23. Bergen SS. Underrepresented minorities in medicine. *JAMA* 2000; 284: 1138–1139.
24. Palepu A, Carr PL, Friedman RH, et al. Minority faculty and academic rank in medicine. *JAMA* 1998; 280: 767–771.
25. Morehouse School of Medicine. *End of Year Faculty Development Evaluation, May, 2007.* Atlanta, GA: Morehouse School of Medicine; 2007.
26. Silberman S, Correa N, Strelnick AH, et al. *Teaching Cases Exploring Cross-Cultural Care.* Bronx, NY: Bronx Center To Reduce and Eliminate Racial and Ethnic Health Disparities; 2007.
27. Hallock VH. *Perceptions of the Professoriate: Anticipatory Socialization of Undergraduate Students from Underrepresented Groups.* Syracuse, NY: Syracuse University; 2003.
28. Leonhardt D. The new affirmative action. *New York Times Magazine.* September 20, 2007; 71–82.
29. US Department of Health and Human Services. An annotated bibliography: evaluation of pipeline development programs designed to increase diversity in the health professions. Available at: http://bhpr.hrsa.gov/healthworkforce/pipelinebibliography.htm. Accessed September 2008.
30. Saha 2006; 8–42.
31. Grumbach K, Munoz C, Coffman J, et al. Strategies for improving the diversity of the health professions. Available at: http://www.calendow.org/uploadedfiles/strategies_for_improving_the_diversity.pdf. Accessed September 2008.

Mount Sinai Journal of Medicine 75:552–555, 2008

EVIDENCE-BASED MEDICINE

The Impact of an Intensive Smoking Cessation Intervention

Deborah Korenstein, MD and Thomas McGinn, MD

Division of General Internal Medicine, Mount Sinai School of Medicine, New York, NY

This is the second in a series of articles in the *Mount Sinai Journal of Medicine* on evidence-based medicine (EBM); this series will emphasize applications of EBM concepts and procedures to clinical practice. The articles will focus on the retrieval, interpretation, and application of evidence related to a current clinical problem. The discussion will be based on the following case.

THE CASE

A 59-year-old man with a history of hypertension and diabetes presents to the emergency department with substernal chest pressure and shortness of breath. The patient has been a 2-pack-per-day smoker for many years. Diagnostic tests reveal an acute myocardial infarction (MI), and the patient is admitted to the cardiac care unit (CCU). Over the next several days, the patient begins to recover from his MI, and the CCU team begins treatment with several different medications to help prevent future MIs and reduce his chance of dying. The CCU intern feels that the patient's smoking is one of his biggest problems and wonders if there are any interventions that could help the patient quit smoking and, more importantly, whether those interventions would prolong his life.

Address Correspondence to:

Deborah Korenstein
Division of General Internal Medicine
Mount Sinai School of Medicine
New York, NY
Email: deborah.korenstein@msnyuhealth.org

FOCUSED CLINICAL QUESTION

To determine whether a smoking cessation intervention would benefit this patient, the intern decides to review the relevant medical literature. She begins by formulating a focused question consisting of 4 elements: the patient (P), the intervention (I), the comparison (C), and the outcome (O). This is often called PICO.[1] The explicit inclusion of these 4 components allows for a more efficient and directed search of the literature and helps to clarify the exact problem of interest. In this case, the question can be framed as follows: in smokers with acute MI (P), does a smoking cessation intervention (I), in comparison with the usual care (C), reduce mortality and improve quit rates (O)?

SEARCH OF THE LITERATURE

The intern decides to go directly to an EBM filtered resource to look for her answer. Filtered resources contain data that have been preselected for quality; some contain expert analysis and interpretation of the data. Commonly used filtered resources include *ACP Journal Club*, which contains brief summaries and reviews of high-quality articles in the literature, and the *Cochrane Database of Systematic Reviews*, which contains systematic reviews related to treatment issues. Filtered databases contain less data than PubMed (the Medline database), but the data that they contain are of uniformly high quality. The intern has only a few minutes to answer her question, and she knows that if she finds a citation in *ACP Journal Club*, it will refer to a high-quality study, so she decides to look there.

Filtered resources generally have less sophisticated search engines than PubMed or other Medline interfaces. Although PubMed will automatically

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20048

© 2008 Mount Sinai School of Medicine

search for terms that are similar to the entered search terms (through automatic MeSH mapping), filtered resources will not, so it is sometimes necessary to redo searches with synonymous terms. In addition, because filtered resources are relatively small databases, it is best to use relatively few and simple search terms to maximize the yield. Performing a search in the *ACP Journal Club* database, the intern types in "smoking cessation myocardial infarction" and clicks on the "therapeutics" category. She gets 14 matches, none of which answers her clinical question. She then tries "smoking cessation mortality" and gets 27 matches. Of these, the sixth one is titled "An Intensive Smoking Cessation Intervention Reduced Mortality in High-Risk Smokers with Cardiovascular Disease."[2] She clicks on that entry to retrieve the article.

Each entry in *ACP Journal Club* is a concise 1-page summary, which is divided into 4 sections (Question, Methods, Main Results, and Conclusions) and is accompanied by a brief editorial. The Main Results section gives the important results in a way that is easy to read and understand, often in table form. In the Methods section, the salient points of validity are addressed, which for treatment or intervention trials include randomization, blinding, and length and completeness of follow-up. The studied population and the details of the intervention are described. *ACP Journal Club* entries provide clinicians with summaries of high-quality articles that are clear, concise, and easy to read.

Validity Criteria for Treatment Articles

Because this article has been reviewed in *ACP Journal Club*, the intern knows that its methodology is essentially valid. The *ACP Journal Club* review refers to a study by Mohiuddin and colleagues[3] titled "Intensive Smoking Cessation Intervention Reduces Mortality in High-Risk Smokers with Cardiovascular Disease," which was published in *Chest* in 2007. The study involved 209 smokers admitted to the CCU with acute coronary syndromes or decompensated heart failure and with significant nicotine dependence, as measured by a Fagerstrom score greater than 7. The major issues[4] of validity are as follows:

- The study was **randomized**. Patients in the intervention group were randomized to receive an intensive smoking cessation intervention or usual care consisting of written material and a 30-minute smoking cessation counseling session. The intervention group received weekly 1-hour counseling sessions for at least 3 months and free individualized therapy with nicotine replacement or bupropion for patients who were felt by physicians to qualify. The intern notes that the unit of randomization in this study was the entire intensive strategy, including both medication and all the components of weekly counseling sessions. Although more patients in the intervention group received drug therapy for smoking cessation, this was not a study of drugs versus no drugs, and it is not clear whether the drugs would have been effective independently of the rest of the intensive strategy.
- The patients in the 2 groups were **similar** with respect to prognostic factors, suggesting that the randomization was generally successful. There were, however, a few differences between the groups. There were slightly more white patients in the intervention group and a trend toward more African Americans in the usual care group. Patients in the intervention group smoked slightly more than those in the control group. Recent studies have shown no significant racial disparity in MI mortality,[5] so the racial imbalance was not likely to have had an impact on the outcome. The fact that intervention patients were heavier smokers would have biased the study in favor of the control group, so the study was unlikely to have overestimated the treatment effect. Patients in the 2 groups were well-balanced with respect to other prognostic variables, such as comorbidities and medications received (other than smoking cessation medications).
- The study was performed and analyzed in an **intention-to-treat** fashion. For example, subjects in the intensive-intervention group were counted as receiving the intervention even if they never attended their counseling or follow-up sessions.
- The study was not **blinded**. Patients and smoking counselors obviously could not be blinded because it was clear what type of care they were receiving or administering. However, a study like this one can blind the outcome assessors, the investigators who determine which patients have reached various study endpoints. This type of blinding is especially important when study endpoints are potentially subjective. For example, if the primary study endpoint were death from cardiovascular disease, unblinded outcome assessors might bias the results through their interpretations of patients' causes of death, whereas if the primary endpoint were overall mortality, there would be no opportunity for interpretation and resultant bias. The study does not state whether outcome assessors were blinded, but the primary endpoints were hospital admissions and overall mortality, both quite objective endpoints, so the lack of blinding is not a major problem.

Results

The results of the study are summarized in the *ACP Journal Club* entry. The intensive-intervention group had higher quit rates than the control group, fewer hospitalizations, and lower mortality. After 2 years of follow-up, 39% of the patients in the intensive-intervention group had quit smoking versus 9% in the control group. Twelve percent of patients in the control group had died after 2 years of follow-up versus 2.8% in the intensive-intervention group. That represents a relative risk reduction for mortality of 77% and a number needed to treat (NNT) of 11 (Table 1). That means that among patients who are similar to the patients in this study, for every 11 patients enrolled in an intensive smoking cessation intervention, we would prevent 1 death over 2 years. Similarly, the rates of hospitalization were 41% in the control group over 2 years and 23% in the intervention group, representing a relative risk reduction of 44% and an NNT of 6. It is noteworthy that 75% of the patients in the intervention group received medication to help them quit smoking versus 17% of the patients in the control group.

Applicability

The intern determines that the study in *Chest* would apply to her patient because the patient would have met the entry criteria.

THE PHYSICIAN'S DECISION

The intern notes that the intensive smoking cessation intervention appears to have an impact on mortality to a very large degree. Even though the study at hand was relatively small, it was well done overall and has reliable results. She wonders how the smoking cessation intervention compares to other interventions in acute MI. Table 2 illustrates the comparative efficacy of different standard treatments during and after an acute MI, with NNTs presented for direct comparison. As the table demonstrates, the NNT of 11 over 2 years for the smoking cessation intervention has an impact on mortality that is comparable, if not superior, to the impact of the standard post-MI treatments used regularly in daily practice.

Table 1. *Mortality Results for the Study.*

EER	CER	ARR (CER − EER)	RRR (ARR/CER)	NNT (1/ARR)
2.8%	12%	9.2%	77%	11

Abbreviations: ARR, absolute risk reduction; CER, control event rate; EER, experimental event rate; NNT, number needed to treat; RRR, relative risk reduction.

Table 2. *Comparison of NNTs for Different Treatments for Acute Myocardial Infarction*[6].

Intervention	Comparison	Time Frame	NNT
Intensive smoking cessation intervention	Usual care	2 years	11
Aspirin	Placebo	Varied	42
Beta blockers	Placebo	Long term	48
PTCA	Thrombolysis	Short term	17
Thrombolysis	Placebo	Short term	53
ACE-I	Placebo	>2 months	42

NOTE: All treatments are for preventing mortality.
Abbreviations: ACE-I, angiotensin-converting enzyme inhibitor; NNT, number needed to treat; PTCA, percutaneous coronary angioplasty.

Is this a landmark study? Landmark studies in the medical literature are those that could change the way that we practice. The impact on mortality of this intensive smoking cessation intervention in CCU smokers is so large that it should change practice. However, as the intern thinks about implementing such an intervention in her patient with an acute MI, she realizes that it would be difficult to accomplish. The patient's insurance provider would not pay for an intensive smoking cessation intervention like this one, and the patients in the study who required medication received it free of charge; this also would not be possible for her patient. If a new pill had such a large mortality effect, it would likely be rapidly approved by the Food and Drug Administration and paid for by insurance companies under pressure from both the pharmaceutical manufacturer and patient advocates. Payment for multifactorial interventions like this one, however, is complex. Convincing insurance carriers to pay for such interventions is challenging, with few natural advocates. In the end, the intern realizes that she could not realistically intervene in this way for her patient, but she decides to counsel him at great length about smoking cessation and to encourage his outpatient physician to do the same on an ongoing basis.

CONCLUSION

This case illustrates several important points:

- Filtered resources such as *ACP Journal Club* can be used to quickly find high-quality literature to answer clinical questions, and they give the reader preanalyzed information that is easy to read and understand.

- It is best to use simple terms to search filtered resources. Try using synonyms if the initial search is not successful.
- In studies of therapies, sometimes the treatment is not a pill but a complex multifactorial intervention. In such cases, we cannot discern which aspect of the intervention affected outcomes; we can see only that the application of the entire strategy had a particular impact.
- The impact of different therapies in a particular patient population can be compared using NNTs.
- Many interventions with behavioral components have been clearly shown to have substantial clinical benefit. However, because of the nature of healthcare financing, these types of evidence-based interventions are rarely implemented in day-to-day practice.

In future articles on EBM, we will discuss other aspects of the evidence cycle. The complete cycle is as follows:

1. A case and a dilemma.
2. A well-formulated clinical question.
3. A search of the literature to answer the question.
4. A critical appraisal of the literature.
5. The incorporation of the patient's values and preferences.
6. Finally, the application of the results to the care of the patient.

Each article in the series will focus on 1 or 2 specific areas, such as methods for searching the medical literature for meaningful data and the procedure for concealment in medical research. Although not meant to be exhaustive, this series will demonstrate how a physician can obtain reliable, evidence-based answers to specific clinical questions of importance to his or her patients.

REFERENCES

1. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 1995; 123: A12–A13.
2. Bentz CJ. An intensive smoking cessation intervention reduced mortality in high-risk smokers with cardiovascular disease. *ACP J Club* 2007; 147: 3.
3. Mohiuddin S, Mooss A, Hunter C, et al. Intensive smoking cessation intervention reduced mortality in high-risk smokers with cardiovascular disease. *Chest* 2007; 131: 446–452.
4. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1993; 270: 2598–2601.
5. Pippins J, Fitzmaurice GM, Haas J. Hospital characteristics and racial disparities in hospital mortality from common medical conditions. *J Natl Med Assoc* 2007; 99: 1030–1036.
6. BMJ clinical evidence. Available at: http://clinicalevidence.bmj.com. Accessed April 1, 2008.

PRISMATIC CASE

Mutation of Growth Hormone Receptor in Children with Idiopathic Short Stature

Frances M. Guevarra, MD[1] and Maria I. New MD[1]

Department of Pediatrics, Division of Adrenal Steroid Disorders, Mount Sinai School of Medicine, New York, NY

Growth hormone (GH) promotes growth by binding to GH receptors and stimulating the production of insulin-like growth factor I (IGF-I). The serum contains a growth hormone–binding protein (GHBP) that is identical to the extracellular domain of the GH receptor.

Growth failure in the majority of short children who do not have GH deficiency is most likely due to a variety of defects. These children with idiopathic short stature (ISS) are otherwise normal and secrete normal levels of GH in response to pharmacologic testing.[1] Children with ISS secrete GH normally but have a defect in the ability of target cells to respond to GH. Such a defect could occur in either the GH receptor or the intracellular mediators of GH signaling. Serum concentrations of GHBP[2] are low in many children with ISS,[3,4] and this suggests that such children may have abnormalities in the gene for the GH receptor. This is supported by the observation that children with low levels of GHBP have lower serum IGF-I concentrations and higher mean 12-hour serum GH concentrations in comparison with children with ISS who have normal levels of GHBP.[5]

Goddard *et al.*[6] reported a prismatic case in which they demonstrated a higher frequency of GH receptor mutations in children with ISS (4 of 14 children) with low levels of GHBP in comparison with normal subjects (0 of 24). An analysis of single-strand conformation polymorphisms and DNA sequencing were both used to identify the mutation in the GH-receptor gene. Children included in the study had standard deviation scores for height below 2.5, GHBP standard deviation scores below −2, IGF-I concentrations below 0, peak GH concentrations with provocative testing above 10 μg, and no systemic illness. All 4 patients found to have GH receptor mutations were treated with recombinant GH. Three of the 4 patients had marginal growth in response to therapy, and this suggests that that they may be partially insensitive to GH because of dysfunctional receptors.

In summary, the identification of this important prismatic case provides a new understanding of potential causes of growth failure in non–GH-deficient children. It also addresses the increased likelihood of these children being partially GH-insensitive and the possible effectiveness of using an alternative form of treatment.

DISCLOSURES

There is nothing to report.

Address Correspondence to:

Maria I. New
Department of Pediatrics
Division of Adrenal Steroid
Disorders
Mount Sinai School of Medicine
Email: maria.new@mssm.edu

REFERENCES

1. Lippe BM, Nakamoto JM. Conventional and unconventional uses of growth hormone. *Recent Prog Horm Res* 1993; 48: 179–235.
2. Carlsson LMS, Rowland AM, Clark RG, et al. Ligand-mediated immunofunctional assay for quantitation of growth hormone-binding protein in human blood. *J Clin Endocrinol Metab* 1991; 73: 1216–1223.
3. Carlsson LMS, Attie KM, Compton PG, et al. Reduced concentration of serum growth hormone-binding

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20039

© 2008 Mount Sinai School of Medicine

protein in children with idiopathic short stature. *J Clin Endocrinol Metab* 1994; 78: 1325–1330.

4. Mauras N, Carlsson LMS, Murphy S, Merimee TJ. Growth hormone-binding protein levels: studies of children with short stature. *Metabolism* 1994; 43: 357–359.
5. Attie KM, Carlsson LMS, Rundle AC, Sherman BM. Evidence for partial growth hormone insensitivity among patients with idiopathic short stature. *J Pediatr* 1995; 127: 244–250.
6. Goddard AD, Covello R, Luoh SH, et al. Mutations of the growth hormone receptor in children with idiopathic short stature. *N Engl J Med* 1995; 333: 1093–1098.

Mount Sinai Journal of Medicine 75:558–566, 2008

Insulin Strategies for Managing Inpatient and Outpatient Hyperglycemia and Diabetes

Etie S. Moghissi, MD, FACE

[1]Department of Medicine, University of California at Los Angeles, Los Angeles, CA
[2]Centinela Hospital, Inglewood, CA

ABSTRACT

Optimal fasting and postprandial glycemic control are essential to limiting microvascular and macrovascular complications associated with diabetes. Recently, stringent control of hyperglycemia in critically ill hospitalized patients with diabetes or acute hyperglycemia has been shown to reduce the risk of morbidity and mortality. This article reviews effective strategies for insulin initiation, titration, and intensification in inpatient and outpatient settings and discusses current treatment strategies when patients are being transitioned from the intensive care unit to general wards and discharged. The development of insulin analogs and premixed insulin analogs has created new options for treating inpatients and outpatients. The more physiologic time-action profiles, improved insulin delivery systems, and standardized protocols for subcutaneous insulin administration and intravenous insulin infusion have improved the safety and convenience of insulin therapy. ***Mt Sinai J Med 75:558–566, 2008.*** © *2008 Mount Sinai School of Medicine*

Key Words: glycemic control, hypoglycemia, inpatient insulin therapy, insulin analogs, outpatient insulin therapy, premixed insulin, treat-to-target.

Diabetes affects nearly 21 million Americans, that is, approximately 7% of the US population.[1] Although most cases are diagnosed in outpatient settings, hospitalization is an important opportunity to identify undiagnosed patients with diabetes and intensify therapy for previously diagnosed but poorly controlled patients. Inpatient hyperglycemia is associated with poor clinical outcomes, increased length of stay, and greater costs in multiple patient populations.[2,3]

Stringent glycemic control decreases microvascular and macrovascular complications as well as morbidity and mortality associated with diabetes.[4–7] Landmark studies (Table 1)[4,6–13] led the American Diabetes Association, American College of Endocrinology, and International Diabetes Federation to establish goals for glycosylated hemoglobin (A1C), fasting plasma glucose (FPG), and postprandial plasma glucose (PPG; Table 2).[2,14–17] Because poor glycemic control may have different effects on outcomes in different populations of hospitalized patients, the utility of these glycemic targets among inpatients has come under scrutiny. Data on the safety of intensive glycemic control and its impact on outcomes among inpatients are evolving. To control hyperglycemia while avoiding hypoglycemia in this patient population, it may be useful to individualize glycemic goals.

Numerous studies have shown that optimal glycemic control decreases morbidity, mortality, and costs in multiple patient populations of critically ill patients.[18–25] In a landmark study, 1548 surgical intensive care unit (ICU) patients were randomly assigned to intensive insulin therapy (target glucose: 80–110 mg/dL) or conventional therapy (target glucose: 180–200 mg/dL).[18] Intensive therapy reduced ICU mortality by 43% and overall hospital mortality by 34% and decreased renal failure, transfusions, and polyneuropathy.[18] Insulin therapy has been shown to prevent kidney injury, accelerate weaning from mechanical ventilation, and accelerate discharge from the medical ICU and the hospital.[20] In cardiac surgery patients, reduced mortality and lower infection rates

Address Correspondence to:

Etie S. Moghissi
Department of Medicine
University of California at Los Angeles
4644 Lincoln Blvd., Suite 409
Marina del Rey, CA 90292
Email: emoghissi@pol.net

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI:10.1002/msj.20071

© 2008 Mount Sinai School of Medicine

Table 1. *Studies Showing That Lowering A1C with Intensive Therapy Reduces Complications in Type 1 and Type 2 Diabetes.*

	DCCT (Type 1)*	Kumamoto (Type 2)†	UKPDS (Type 2)‡	DCCT/EDIC (Type 1)§
Target A1C reduction	9% → 7%	9% → 7%	8% → 7%	9% → 8%
Complication				
Retinopathy	↓ 63%	↓ 69%	↓ 17%–21%	↓ 72%–87%
Nephropathy	↓ 54%	↓ 70%	↓ 24%–33%	↓ 53%–92%
Neuropathy	↓ 60%	–	–	–
Macrovascular disease	↓ 41%	–	↓ 16%¶	↓ 42%–57%

Abbreviations: A1C, glycosylated hemoglobin; DCCT, Diabetes Control and Complications Trial: EDIC, Epidemiology of Diabetes Interventions and Complications Study; UKPDS, United Kingdom Prospective Diabetes Study.

[*] Data were taken from references 4 and 8.

[†] Data were taken from reference 9.

[‡] Data were taken from references 6, 10, 11, and 12.

[§] Data were taken from references 7 and 13.

[¶] Myocardial infarction.

Table 2. *ADA, ACE/AACE, and IDF Glycemic Goals.*

Parameter	ADA Guidelines*	ACE/AACE Guidelines†	IDF Guidelines‡
Outpatients			
A1C	< 7%§	≤ 6.5%	< 6.5%
Fasting plasma glucose	< 90 –130 mg/dL	<110 mg/dL	<110 mg/dL
Postprandial plasma glucose	< 180 mg/dL	<140 mg/dL	<145 mg/dL
Inpatients			
Intensive care unit	As close to 110 mg/dL as possible, <180 mg/dL	110 mg/dL¶	–
Noncritical care			
Preprandial glucose	90-130 mg/dL	110 mg/dL¶	–
Maximal glucose	<180 mg/dL	180 mg/dL¶	–

Abbreviations: A1C, glycosylated hemoglobin; AACE, American Association for Clinical Endocrinologists; ACE, American College of Endocrinology; ADA, American Diabetes Association; IDF, International Diabetes Federation.

[*] Data were taken from reference 14.

[†] Data were taken from references 2, 15, and 16.

[‡] Data were taken from reference 17.

[§] As close to normal (< 6.0%) as possible without hypoglycemia in individual patients.

[¶] Upper limits for glycemic targets.

occurred when euglycemia was maintained for 3 days postoperatively.[2] Insulin therapy may reduce morbidity and mortality by suppressing proinflammatory cytokines and prothrombotic growth factors,[26,27] promoting vasodilation,[28] and reducing infection risk due to impaired immunity.[22]

Inpatient glycemic targets have been established on the basis of compelling evidence that optimal glycemic control can improve clinical outcome. According to the American College of Endocrinology and the American Association for Clinical Endocrinologists, glycemic targets are < 110 mg/dL in the ICU and < 110 mg/dL before meals and < 180 mg/dL for peak glucose in patients in noncritical care units.[2] The American Diabetes Association recommends that blood glucose levels be kept as close to 110 mg/dL as possible in critically ill patients and generally < 180 mg/dL. In noncritically ill patients, premeal blood glucose should be kept as close to 90 to 130 mg/dL as possible, and postprandial glucose should be < 180 mg/dL.[14] Glycemic targets should be individualized for elderly patients or those with severe comorbidities (liver or renal disease).

Despite the importance of optimal glycemic control and new therapies, most patients are not meeting the goals. Less than half of all people with diabetes achieve A1C < 7%.[29] In hospitals, glycemic control is also suboptimal because of reluctance to use insulin intensively, tolerance of moderate hyperglycemia, reliance on the nonphysiologic insulin replacement sliding-scale strategy, fear of hypoglycemia, and the old belief that stress hyperglycemia is benign in patients without diabetes.[30–32] In fact, a recent national survey reported that 18% to 41% of inpatients with diabetes experienced hyperglycemia (defined as blood glucose levels > 200 mg/dL) for 3 consecutive days.[33]

INITIATING INSULIN THERAPY

A major impediment is the reluctance of patients and physicians to initiate insulin (the most effective antidiabetic therapy when dosed appropriately)[34] when oral antidiabetic drugs (OADs), either alone or in combination, are no longer effective because of progressive β-cell failure.[35] As a result, in clinical practice, insulin initiation is often delayed and is considered only after long periods of inadequate control.[36,37]

The American College of Endocrinology recommends initiating insulin as initial therapy if A1C is > 10% in patients not using pharmacotherapy or if A1C is > 8.5% in patients treated with antidiabetic agents.[38] The American Diabetes Association and European Association for the Study of Diabetes recommend initiating insulin when A1C is > 10%, FPG is > 250 mg/dL, or glycemic targets are not met with metformin and a second agent.[39] Insulin should be considered in the majority of hospitalized patients to achieve inpatient glycemic control.

OPTIONS FOR INSULIN REPLACEMENT

In individuals without diabetes, low levels of basal insulin limit hepatic glucose output between meals. Insulin levels peak 15 to 45 minutes after meals and return to basal levels 60 to 180 minutes later.[40–42] The ideal basal insulin would have a slow onset of action and a duration of action long enough to allow once daily dosing with no distinct peak. The ideal prandial insulin would have a rapid onset of action to control the initial rise in postprandial blood glucose and a short duration of action to prevent postprandial hypoglycemia.

Human Insulin Preparations

Neutral protamine Hagedorn (NPH) insulin has a relatively short duration of action that requires twice daily dosing for 24-hour coverage and a distinct and unpredictable peak.[43–45] Its variable onset (and short duration) of action increases the risk of hypoglycemia, particularly at night.[44] Nonetheless, NPH insulin may be used in patients on tube feeds (when dosed every 6–8 hours) and in patients using steroids (when dosed in the morning).

Use of regular human insulin for prandial coverage has significant drawbacks, including a delayed onset of action that necessitates injection at least 30 minutes before eating,[43] which can be inconvenient, particularly when mealtimes are unpredictable. The long duration of action also increases the risk of late postprandial hypoglycemia.[46]

Premixed human insulin therapies have the same limitations as their individual components: the unpredictability and distinct peak of NPH and the slow onset and long duration of action of regular human insulin.[47] They also must be dosed 30 minutes before meals.

Insulin Analog Preparations

The deficiencies of human insulins led to the development of insulin analogs. Rapid-acting insulin analogs–insulin lispro, aspart, and glulisine–have a more rapid onset, greater peak, and shorter duration of action than regular human insulin.[48] Therefore, rapid-acting insulin analogs can be administered within 15 minutes of meal initiation. These improvements also reduce the risk of hypoglycemia in comparison with regular human insulin.[48–50] Although regular human insulin is considerably cheaper than insulin analogs, its inconvenient need to be administered 30 minutes before meals makes it difficult to use in busy hospital settings.[51]

To address the limitations of basal human insulins, long-acting insulin analogs with prolonged activity, a relatively flat time-action profile, and more consistent absorption were developed.[48] Two basal insulin analogs are now available: insulin detemir and insulin glargine. They have more predictable effects[52,53] and are associated with considerably less hypoglycemia.[54,55] Insulin detemir is also associated with less weight gain.[52,55] In addition, long-acting insulin analogs can be dosed once daily in type 2 diabetes.[52,54] In type 1 diabetes, insulin detemir can be dosed once or twice daily.[56]

For patients who need basal and prandial coverage and wish to minimize the multiple injections required with basal/bolus regimens, a premixed insulin analog such as biphasic insulin aspart 70/30 (BIAsp 30), insulin lispro 75/25, or insulin lispro 50/50 can be used. Premixed insulin analogs contain 25% to 50% of a rapid-acting analog for meal coverage and 50% to 75% of a protaminated analog for basal needs. In contrast to premixed human insulins, premixed insulin analogs enable flexible dosing and can substantially reduce PPG excursions.[47]

STRATEGIES FOR ATTAINING GLUCOSE TARGETS IN HOSPITALIZED PATIENTS

Insulin given either intravenously as a continuous infusion or subcutaneously is the most effective agent for controlling hyperglycemia in hospitalized patients.[2,3,57] All inpatients with type 1 diabetes should receive intravenous or subcutaneous insulin and should never be left without insulin coverage.

Although regular human insulin is primarily used for continuous insulin infusion (CII),[2] insulin aspart can also be administered intravenously.[58] Table 3 highlights indications for CII. Although CII improves glycemic control and reduces hypoglycemia in various inpatient settings,[3,31,57,59–62] enough glucose (5–10 g/hour) must be provided to prevent starvation ketosis and hypoglycemia. Patients on parenteral nutrition should be dosed with regular insulin, and the initial dose can be calculated by the division of the amount of dextrose in a 24-hour intravenous bag by 10. The dose of insulin should usually be gradually titrated in increments of 5 to 10 U/L until glycemic control is achieved.[51] Routine bedside glucose testing is needed to ensure safety.

Traditionally, the insulin sliding-scale protocol has been used to manage inpatient hyperglycemia.[63] Sliding-scale insulin orders typically provide a predetermined amount of regular human insulin, beginning with 2 units for blood glucose > 150 mg/dL and increasing by 2 units for every 50 mg/dL above this target.[64] However, sliding-scale insulin has many limitations: it treats hyperglycemia rather than preventing it[65] and does not account for insulin resistance, meals, weight, illness, or basal insulin requirements.[66] Therefore, sliding-scale insulin is ineffective and potentially dangerous.[31,63,65,67,68]

Various CII protocols,[3] developed to address the limitations of sliding-scale insulin, have been proven effective in inpatients and have been published.[21,61,62,69] Protocols that use dynamic scales, which incorporate the rate of change in glucose into dose adjustments, are typically the most effective.[51] Dynamic protocols take into account several factors, including the patient's current and previous blood glucose levels (therefore, the rate of change in blood glucose levels) and the current insulin infusion rate.

The ideal CII protocol is easily ordered and implemented and is effective and safe. Successful implementation requires a multidisciplinary effort with support from administration, nursing, and pharmacy staff.

Table 3. *Indications for Intravenous Insulin Therapy.*

Prolonged NPO in type 1 diabetes
Critical illness
Preoperative period
After organ transplantation
Diabetic ketoacidosis
Total parenteral nutrition therapy
Labor and delivery
Myocardial infarction
Other illnesses requiring prompt glucose control
Use of high doses of corticosteroids[57]

Abbreviation: NPO, nil per os.

EFFECTIVE SUBCUTANEOUS INSULIN THERAPY

If CII is not required, the use of a subcutaneous insulin regimen that meets basal, nutritional, and supplemental needs is optimal. Nutritional insulin is that needed to cover any intravenous glucose plus intravenous or enteral alimentation and meals. If the patient is eating and not receiving any other sources of calories, nutritional insulin would be prandial insulin. Inpatients often require higher insulin doses because of stress related to critical illness. Subcutaneous insulin can be given as a scheduled dose (basal plus nutritional) and a supplemental (correction) dose if glycemic levels remain above target. The supplemental algorithm should not be confused with sliding-scale insulin, which traditionally has been used with no scheduled dose. Appropriate diabetes care requires timely coordination of point-of-care blood glucose testing, insulin administration, and delivery of meals.

Insulin analogs can minimize hypoglycemia and increase flexibility in comparison with human insulin therapies. In outpatients, insulin analogs result in a lower risk of hypoglycemia. Rapid-acting analogs also provide better postprandial control. Although not studied, it is likely that they would be effective and safe in the hospital setting and would be a preferred type of insulin for inpatients. Because food intake is often unpredictable among inpatients, a prudent approach to using prandial insulin therapies is to administer the insulin only when the meal tray is in front of the patient.[51] When dietary intake is uncertain, clinicians should be conservative in dosing prandial insulin or administer the rapid-acting insulin analog immediately after the meal when the amount of food consumed is known. For supplemental insulin coverage, the rapid-acting analogs are preferred.

Hypoglycemia is a major barrier to aggressive glycemic control in the hospital and outpatient settings, but it is predictable and preventable. Factors that increase the risk of inpatient hypoglycemia include inadequate glucose monitoring; lack of coordination between meals, transportation, and nursing; indecipherable orders; and unsafe work environments. Clear communication between nursing and ancillary services, frequent glucose monitoring, clear processes for insulin orders, and standard hypoglycemia protocols are critical to prevent hypoglycemia. Incorrect dosing and timing of insulin administration, missed meals, and using a basal insulin that peaks during sleep increase the risk of hypoglycemia among outpatients.

Recent studies have evaluated the use of insulin analogs among inpatients. In 1 recent prospective, randomized study, the efficacy of a basal/bolus insulin regimen using insulin glargine and insulin glulisine was compared to that of a standard sliding-scale protocol with regular insulin in 130 noncritically ill insulin-naïve inpatients.[70] In this study, 66% of patients who received insulin glargine/glulisine achieved a blood glucose target of < 140 mg/dL versus only 38% of patients who received sliding-scale insulin. Rates of hypoglycemia were similar between groups. These results were confirmed in a recent study comparing insulin glargine to 6-hour sliding-scale regular insulin in 81 patients with postoperative blood glucose levels greater than 144 mg/dL following gastric bypass surgery.[71] In this study, mean blood glucose levels were 154 mg/dL with sliding-scale insulin and 134 mg/dL with insulin glargine ($P < 0.01$). Hypoglycemia in both groups was infrequent. These results indicate that treatment with a basal insulin analog, with or without rapid-acting insulin, resulted in a significant improvement in glycemic control in comparison with sliding-scale insulin without an increase in the risk of hypoglycemia.[70,71]

Detection of inpatient hyperglycemia is an opportunity to improve long-term care. An elevated A1C indicates hyperglycemia before admission and warrants re-evaluation and modification of long-term therapy. In those without a history of diabetes, differentiation between hospital-related hyperglycemia and undiagnosed diabetes is necessary. These patients require follow-up testing (fasting or 2-hour oral glucose testing) for evaluation.

INSULIN USE IN PREOPERATIVE PATIENTS

Treatment recommendations for a preoperative patient with diabetes depend on the patient's type of diabetes, surgery, preoperative treatment regimen, and state of metabolic control.[72] For all patients, oral agents should be withheld on the day of the surgery.[72] All patients with type 1 or type 2 diabetes treated with insulin who have fair metabolic control (FPG < 180 mg/dL) may receive their usual basal insulin analog dose or half their NPH dose the morning of the surgery. Individuals with poor glycemic control (FPG > 180 mg/dL) should be considered for CII. Prandial rapid-acting or short-acting insulin therapy should be withheld until meals are tolerated. In patients on insulin undergoing major surgery, CII should be initiated before surgery and continued during the perioperative period.

TRANSITIONING FROM INTRAVENOUS INSULIN TO SUBCUTANEOUS INSULIN AND FROM INPATIENT INSULIN THERAPY TO OUTPATIENT INSULIN THERAPY

Although acute hyperglycemia may normalize in some inpatients before discharge, some patients diagnosed with diabetes may need to transition to subcutaneous insulin therapy. In order to maintain glycemic control in inpatients transitioning to subcutaneous insulin, short-acting or rapid-acting insulin should be injected 1 to 2 hours before discontinuation of CII. Basal and prandial insulin doses must be individualized. To establish the 24-hour insulin requirement, one can extrapolate from the average CII dose required over the previous 4 to 6 hours (if stable) and give half as basal insulin and the other half as prandial insulin divided by the number of meals. Some patients with type 2 diabetes who have mild or moderate hyperglycemia may initiate therapy with basal insulin alone, particularly if discharge is imminent.[51] The dose should be adjusted according to the patient's stress level, food intake, intravenous or enteral alimentation, weight, insulin sensitivity, and concomitant medication use (eg, steroids). If a premixed insulin analog is being used, half can be given with breakfast, and half can be given with supper. If basal insulin alone is used, the total dose should be administered once daily in the evening. Using newer insulin delivery systems, such as pen devices, may improve treatment convenience and acceptance.

Regardless of the regimen used, clinicians should establish realistic glycemic targets, treat comorbidities, and recommend lifestyle changes. Patients should also be evaluated by a dietitian and a certified diabetes educator while the attending physician provides necessary details of the treatment plan to the patient's primary care physician. In patients who require insulin therapy, initiating treat-to-target regimens that involve frequent monitoring of glycemic control and dosage adjustments may be most effective in helping patients achieve and maintain their glycemic goals.

THE TREAT-TO-TARGET PARADIGM

Treat-to-target paradigms are relatively new approaches that employ frequent monitoring of glycemic targets and appropriate adjustment of therapy. Pharmacologic agents are introduced and intensified, alone or in combination, as needed to reach glycemic goals. Several key studies evaluating the treat-to-target approach using the 2 main approaches to initiating insulin therapy are reviewed here.[55,73–77]

Long-Acting Insulin Analogs

The first treat-to-target study compared glargine to NPH, both dosed once daily, in 756 patients with type 2 diabetes inadequately controlled with OADs.[73] Insulin was added to OADs and titrated weekly to a target FPG ($\leq$ 100 mg/dL). More patients using glargine than patients using NPH achieved an A1C $\leq$ 7% without nocturnal hypoglycemia (33.2% versus 26.7%, $P < 0.05$). Overall hypoglycemia and nocturnal hypoglycemia rates were lower with glargine.

Similar results occurred with insulin detemir. In 1 study, 476 patients inadequately controlled on OADs were randomly assigned to either detemir or NPH for 26 weeks. Insulin was titrated weekly to achieve FPG and presupper glucose levels $\leq$ 108 mg/dL. In comparison with NPH, nocturnal hypoglycemia was 55% lower ($P < 0.001$) and overall hypoglycemia was 47% lower ($P < 0.001$) with detemir.[55] Less weight gain (1.2 versus 2.8 kg, $P < 0.001$) and glucose variability occurred with detemir compared with NPH, and these findings have been confirmed in other studies.[52,55,73,74]

Using Premixed Insulin Preparations to Control Both Fasting Plasma Glucose and Postprandial Plasma Glucose

Premixed insulins enable FPG and PPG to be covered in 1 injection. In a treat-to-target trial, 233 patients with inadequate glycemic control on OADs received either BIAsp 30 twice daily or glargine once daily. All patients used metformin, and about one-third also used pioglitazone. Mean A1C values after 28 weeks were 6.9% and 7.4% for BIAsp 30–treated and glargine-treated patients, respectively ($P < 0.01$); 66% reached an A1C <7% with BIAsp 30 versus 40% with glargine ($P < 0.01$). More episodes of hypoglycemia occurred with BIAsp 30, but none were serious or required discontinuation of treatment.[75]

In a 32-week, randomized, crossover study with 97 patients, those on lispro 75/25 twice daily were more likely to reach an A1C $\leq$ 7% (30% versus 12%, $P = 0.002$), had lower end-study mean A1C (7.5% versus 8.1%, $P < 0.001$) and PPG measurements (173 versus 202, $P < 0.001$), gained less weight, and had less nocturnal hypoglycemia than those using glargine every day. Those using glargine had a significantly lower FPG (133 versus 142 mg/dL, $P = 0.007$).[76]

The 1-2-3 Study used a treat-to-target algorithm in which BIAsp 30 every day was added to OADs at supper and titrated in 3-unit increments based on fasting blood glucose. At week 16, an additional prebreakfast injection of BIAsp 30 was added if A1C was >6.5%, whereas patients who remained above this goal at week 32 added a third injection at lunch. A1C levels $\leq$ 6.5% were achieved by 21%, 52%, and 60% of patients taking 1, 2, and 3 injections, respectively. A1C levels < 7% were achieved by 41%, 70%, and 77% of patients, respectively.[77]

SUMMARY AND CONCLUSIONS

There is overwhelming evidence that hyperglycemia is an independent risk factor for poor outcomes in the hospital and outpatient settings and that optimal glycemic control can improve clinical outcomes. Insulin therapy can be tailored to meet physiologic requirements, and targeted glucose goals can be achieved while the risk of hypoglycemia is minimized, particularly with insulin analogs.

The challenge now facing institutions is safe and effective implementation of the new recommendations. Standardized protocols developed by multidisciplinary teams (including physicians, diabetes educators, and nurses) are necessary to ensure safety and efficacy. Using this collaborative approach can improve clinical outcomes and reduce length of stay. Discharge planning should be initiated early to evaluate community resources and arrange proper follow-up care.

In outpatients inadequately controlled on OADs, basal or premixed insulins can be added.[78] Insulin glargine and insulin detemir are preferable to NPH because they lower FPG and A1C like NPH but do so with a substantially lower risk of hypoglycemia. Premixed insulin analogs also provide a useful therapeutic option for patients requiring both basal and prandial insulin.[42,47]

Insulin analogs that more closely mimic endogenous insulin secretion are now available. They are more physiological, are conveniently dosed, and reduce the potential for hypoglycemia. Therefore, improved glycemic control and subsequent reductions in morbidity and mortality may be within sight for inpatients with acute hyperglycemia and patients with diabetes.

ACKNOWLEDGMENTS

The author thanks Novo Nordisk for funding that helped support the preparation of this article and Nicole Cooper for writing and editorial support.

DISCLOSURES

Potential conflict of interest: Etie S. Moghissi has been a speaker for Abbott, Amylin, Eli Lilly, Novartis, Novo Nordisk, and Pfizer for the past 3 years. Novo Nordisk provided funding that helped support the preparation of this article.

REFERENCES

1. Centers for Disease Control. National diabetes fact sheet. United States, 2005: general information. Available at: www.cdc.gov/diabetes/pubs/pdf/ndfs_2005.pdf. Accessed August 2007.
2. Garber AJ, Moghissi ES, Bransome ED Jr, et al. American College of Endocrinology position statement on inpatient diabetes and metabolic control. *Endocr Pract* 2004; 10: 77–82.
3. Moghissi ES, Hirsch IB. Hospital management of diabetes. *Endocrinol Metab Clin North Am* 2005; 34: 99–116.
4. Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* 1993; 329: 977–986.
5. Ohkubo Y, Kishikawa H, Araki E, et al. Intensive insulin therapy prevents the progression of diabetic microvascular complications in Japanese patients with non-insulin-dependent diabetes mellitus: a randomized prospective 6-year study. *Diabetes Res Clin Pract* 1995; 28: 103–117.
6. United Kingdom Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 1998; 352: 837–853.
7. DCCT/EDIC Research Group. Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. *N Engl J Med* 2005; 353: 2643–2653.
8. DCCT Research Group. Effect of intensive therapy on the development and progression of diabetic nephropathy in the Diabetes Control and Complications Trial. *Kidney Int* 1995; 47: 1703–1720.
9. Shichiri M, Kishikawa H, Ohkubo Y, Wake N. Long-term results of the Kumamoto study on optimal diabetes control in type 2 diabetic patients. *Diabetes Care* 2000; 23(suppl 2): B21–B29.
10. UK Prospective Diabetes Study Group. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). *Lancet* 1998; 352: 854–865.
11. UK Prospective Diabetes Study Group. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. *BMJ* 1998; 317: 703–713.
12. Wright A, Burden AC, Paisey RB, et al. Sulfonylurea inadequacy: efficacy of addition of insulin over 6 years in patients with type 2 diabetes in the U.K. Prospective Diabetes Study (UKPDS 57). *Diabetes Care* 2002; 25: 330–336.
13. EDIC Research Group. Epidemiology of Diabetes Interventions and Complications (EDIC): design, implementation, and preliminary results of a long-term follow-up of the diabetes control and complications trial cohort. *Diabetes Care* 1999; 22: 99–111.
14. American Diabetes Association. Standards of medical care in diabetes–2006. *Diabetes Care* 2006; 29(suppl 1): S4–S42.
15. American Association of Clinical Endocrinologists. The American Association of Clinical Endocrinologists medical guidelines for the management of diabetes mellitus: the AACE system of intensive diabetes self-management–2002 update. *Endocr Pract* 2002; 8(suppl 1): 40–82.
16. Lebovitz HE, Austin MM, Blonde L, et al. ACE/AACE consensus conference on the implementation of outpatient management of diabetes mellitus: consensus conference recommendations. *Endocr Pract* 2006; 12(suppl 1): 6–12.
17. IDF Clinical Guidelines Task Force. Global guideline for type 2 diabetes. Available at: http://www.idf.org/webdata/docs/IDF%20GGT2D.pdf Accessed August 2007.
18. Van den Berghe G, Wouters P, Weekers F, et al. Intensive insulin therapy in the critically ill patients. *N Engl J Med* 2001; 345: 1359–1367.
19. Pittas AG, Siegel RD, Lau J. Insulin therapy for critically ill hospitalized patients: a meta-analysis of randomized controlled trials. *Arch Intern Med* 2004; 164: 2005–2011.

20. Van den Berghe G, Wilmer A, Hermans G, et al. Intensive insulin therapy in the medical ICU. *N Engl J Med* 2006; 354: 449–461.
21. Furnary AP, Wu Y, Bookin SO. Effect of hyperglycemia and continuous intravenous insulin infusions on outcomes of cardiac surgical procedures: the Portland Diabetic Project. *Endocr Pract* 2004; 10(suppl 2): 21–33.
22. Furnary AP, Zerr KJ, Grunkemeier GL, Starr A. Continuous intravenous insulin infusion reduces the incidence of deep sternal wound infection in diabetic patients after cardiac surgical procedures. *Ann Thorac Surg* 1999; 67: 352–360.
23. Furnary AP, Gao G, Grunkemeier GL, et al. Continuous insulin infusion reduces mortality in patients with diabetes undergoing coronary artery bypass grafting. *J Thorac Cardiovasc Surg* 2003; 125: 1007–1021.
24. Capes SE, Hunt D, Malmberg K, Gerstein HC. Stress hyperglycaemia and increased risk of death after myocardial infarction in patients with and without diabetes: a systematic overview. *Lancet* 2000; 355: 773–778.
25. Van den Berghe G, Wouters PJ, Bouillon R, et al. Outcome benefit of intensive insulin therapy in the critically ill: insulin dose versus glycemic control. *Crit Care Med* 2003; 31: 359–366.
26. Dandona P, Aljada A, Bandyopadhyay A. The potential therapeutic role of insulin in acute myocardial infarction in patients admitted to intensive care and in those with unspecified hyperglycemia. *Diabetes Care* 2003; 26: 516–519.
27. Dandona P, Aljada A, Mohanty P, et al. Insulin inhibits intranuclear nuclear factor kappaB and stimulates IkappaB in mononuclear cells in obese subjects: evidence for an anti-inflammatory effect? *J Clin Endocrinol Metab* 2001; 86: 3257–3265.
28. Steinberg HO, Brechtel G, Johnson A, et al. Insulin-mediated skeletal muscle vasodilation is nitric oxide dependent. A novel action of insulin to increase nitric oxide release. *J Clin Invest* 1994; 94: 1172–1179.
29. Saaddine JB, Cadwell B, Gregg EW, et al. Improvements in diabetes processes of care and intermediate outcomes: United States, 1988-2002. *Ann Intern Med.* 2006; 144: 465–474.
30. Abourizk NN, Vora CK, Verma PK. Inpatient diabetology. The new frontier. *J Gen Intern Med* 2004; 19: 466–471.
31. Clement S, Braithwaite SS, Magee MF, et al. Management of diabetes and hyperglycemia in hospitals. *Diabetes Care* 2004; 27: 553–591.
32. Ellger B, Debaveye Y, Vanhorebeek I, et al. Survival benefits of intensive insulin therapy in critical illness: impact of maintaining normoglycemia versus glycemia-independent actions of insulin. *Diabetes* 2006; 55: 1096–1105.
33. Wexler DJ, Meigs JB, Cagliero E, et al. Prevalence of hyper- and hypoglycemia among inpatients with diabetes: a national survey of 44 U.S. hospitals. *Diabetes Care* 2007; 30: 367–369.
34. Riddle M. Introduction. *Am J Med* 2000; 108(suppl 6A): 1.
35. Kirkman MS, Williams SR, Caffrey HH, Marrero DG. Impact of a program to improve adherence to diabetes guidelines by primary care physicians. *Diabetes Care* 2002; 25: 1946–1951.
36. Raskin P. Can glycemic targets be achieved–in particular with two daily injections of a mix of intermediate- and short-acting insulin? *Endocr Pract* 2006; 12(suppl 1): 52–54.
37. Peyrot M, Rubin RR, Lauritzen T, et al. Resistance to insulin therapy among patients and providers: results of the cross-national Diabetes Attitudes, Wishes, and Needs (DAWN) study. *Diabetes Care* 2005; 28: 2673–2679.
38. Davidson JA, Blonde L, Jellinger PS, et al. Road map for the prevention and treatment of type 2 diabetes. ACE/AACE Diabetes Road Map Task Force. Available at: www.aace.com/meetings/consensus/odimplementation/roadmap.pdf. Accessed August 2007.
39. Nathan DM, Buse JB, Davidson MB, et al. Management of hyperglycaemia in type 2 diabetes: a consensus algorithm for the initiation and adjustment of therapy. A consensus statement from the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetologia* 2006; 49: 1711–1721.
40. Polonsky KS, Given BD, van Cauter E. Twenty-four-hour profiles and pulsatile patterns of insulin secretion in normal and obese subjects. *J Clin Invest* 1988; 81: 442–448.
41. Lindholm A, McEwen J, Riis AP. Improved postprandial glycemic control with insulin aspart. A randomized double-blind cross-over trial in type 1 diabetes. *Diabetes Care* 1999; 22: 801–805.
42. Rolla AR, Rakel RE. Practical approaches to insulin therapy for type 2 diabetes mellitus with premixed insulin analogues. *Clin Ther* 2005; 27: 1113–1125.
43. DeWitt DE, Hirsch IB. Outpatient insulin therapy in type 1 and type 2 diabetes mellitus: scientific review. *JAMA* 2003; 289: 2254–2264.
44. Kurtzhals P. How to achieve a predictable basal insulin? *Diabetes Metab* 2005; 31: 4S25–4S33.
45. Heinemann L, Richter B. Clinical pharmacology of human insulin. *Diabetes Care* 1993; 16(suppl 3): 90–100.
46. Dimitriadis GD, Gerich JE. Importance of timing of preprandial subcutaneous insulin administration in the management of diabetes mellitus. *Diabetes Care* 1983; 6: 374–377.
47. Garber AJ. Premixed insulin analogues for the treatment of diabetes mellitus. *Drugs* 2006; 66: 31–49.
48. Hirsch IB. Insulin analogues. *N Engl J Med* 2005; 352: 174–183.
49. Brunelle BL, Llewelyn J, Anderson JH Jr, et al. Meta-analysis of the effect of insulin lispro on severe hypoglycemia in patients with type 1 diabetes. *Diabetes Care* 1998; 21: 1726–1731.
50. Home PD, Lindholm A, Hylleberg B, Round P. Improved glycemic control with insulin aspart: a multicenter randomized double-blind crossover trial in type 1 diabetic patients. *Diabetes Care* 1998; 21: 1904–1909.
51. Inzucchi SE. Clinical practice. Management of hyperglycemia in the hospital setting. *N Engl J Med* 2006; 355: 1903–1911.
52. Haak T, Tiengo A, Draeger E, et al. Lower within-subject variability of fasting blood glucose and reduced weight gain with insulin detemir compared to NPH insulin in patients with type 2 diabetes. *Diabetes Obes Metab* 2005; 7: 56–64.
53. Raskin P, Klaff L, Bergenstal R, et al. A 16-week comparison of the novel insulin analog insulin glargine

(HOE 901) and NPH human insulin used with insulin lispro in patients with type 1 diabetes. *Diabetes Care* 2000; 23: 1666–1671.

54. Rosenstock J, Schwartz SL, Clark CM Jr. et al. Basal insulin therapy in type 2 diabetes: 28-week comparison of insulin glargine (HOE 901) and NPH insulin. *Diabetes Care* 2001; 24: 631–636.
55. Hermansen K, Davies M, Derezinski T, et al. A 26-week, randomized, parallel, treat-to-target trial comparing insulin detemir with NPH insulin as add-on therapy to oral glucose-lowering drugs in insulin-naive people with type 2 diabetes. *Diabetes Care* 2006; 29: 1269–1274.
56. *Levemir® (Insulin Detemir [RDNA Origin] Injection).* Princeton, NJ: Novo Nordisk; 2005.
57. Trence DL, Kelly JL, Hirsch IB. The rationale and management of hyperglycemia for in-patients with cardiovascular disease: time for change. *J Clin Endocrinol Metab* 2003; 88: 2430–2437.
58. *NovoLog® Mix 70/30 (70% Insulin Aspart Protamine Suspension and 30% Insulin Aspart Injection [RDNA Origin]).* Princeton, NJ: Novo Nordisk; 2005.
59. Quevedo SF, Sullivan E, Kington R, Rogers W. Improving diabetes care in the hospital using guideline-directed orders. *Diabetes Spectr* 2001; 14: 226–233.
60. Brown G, Dodek P. Intravenous insulin nomogram improves blood glucose control in the critically ill. *Crit Care Med* 2001; 29: 1714–1719.
61. Markovitz LJ, Wiechmann RJ, Harris N, et al. Description and evaluation of a glycemic management protocol for patients with diabetes undergoing heart surgery. *Endocr Pract* 2002; 8: 10–18.
62. Ku SY, Sayre CA, Hirsch IB, Kelly JL. New insulin infusion protocol improves blood glucose control in hospitalized patients without increasing hypoglycemia. *Jt Comm J Qual Patient Saf* 2005; 31: 141–147.
63. Umpierrez GE, Palacio A, Smiley D. Sliding scale insulin use: myth or insanity? *Am J Med* 2007; 120: 563–567.
64. Baldwin D, Villanueva G, McNutt R, Bhatnagar S. Eliminating inpatient sliding-scale insulin: a reeducation project with medical house staff. *Diabetes Care* 2005; 28: 1008–1011.
65. Gill G, MacFarlane I. Are sliding-scale insulin regimens a recipe for diabetic instability? *Lancet* 1997; 349: 1555.
66. Lorber DL. Sliding scale insulin. *Diabetes Care* 2001; 24: 2011–2012.
67. Shagan BP. Does anyone here know how to make insulin work backwards? Why sliding-scale insulin coverage doesn't work. *Pract Diabetol* 1990; 9: 1–4.
68. Queale WS, Seidler AJ, Brancati FL. Glycemic control and sliding scale insulin use in medical inpatients with diabetes mellitus. *Arch Intern Med* 1997; 157: 545–552.
69. American Association of Clinical Endocrinologists and American Diabetes Association. Inpatient diabetes and glycemic control: a call to action conference. Available at: http://www.aace.com/meetings/consensus/IIDC/IDGC0207.pdf. Accessed August 2007.
70. Umpierrez GE, Smiley D, Zisman A, et al. Randomized study of basal bolus insulin therapy in the inpatient management of patients with type 2 diabetes (RABBIT 2 Trial). *Diabetes Care* 2007; 30: 2181–2186.
71. Datta S, Qaadir A, Villanueva G, Baldwin D. Once-daily insulin glargine versus 6-hour sliding scale regular insulin for control of hyperglycemia after a bariatric surgical procedure: a randomized clinical trial. *Endocr Pract* 2007; 13: 225–231.
72. Smiley DD, Umpierrez GE. Perioperative glucose control in the diabetic or nondiabetic patient. *South Med J* 2006; 99: 580–589.
73. Riddle MC, Rosenstock J, Gerich J. The treat-to-target trial: randomized addition of glargine or human NPH insulin to oral therapy of type 2 diabetic patients. *Diabetes Care* 2003; 26: 3080–3086.
74. Yki-Jarvinen H, Kauppinen-Makelin R, Tiikkainen M, et al. Insulin glargine or NPH combined with metformin in type 2 diabetes: the LANMET study. *Diabetologia* 2006; 49: 442–451.
75. Raskin P, Allen E, Hollander P, et al. Initiating insulin therapy in type 2 diabetes: a comparison of biphasic and basal insulin analogs. *Diabetes Care* 2005; 28: 260–265.
76. Malone JK, Bai S, Campaigne BN, et al. Twice-daily pre-mixed insulin rather than basal insulin therapy alone results in better overall glycaemic control in patients with type 2 diabetes. *Diabet Med* 2005; 22: 374–381.
77. Garber AJ, Wahlen J, Wahl T, et al. Attainment of glycaemic goals in type 2 diabetes with once-, twice-, or thrice-daily dosing with biphasic insulin aspart 70/30 (the 1-2-3 study). *Diabetes Obes Metab* 2006; 8: 58–66.
78. Rosenstock J. Basal insulin supplementation in type 2 diabetes; refining the tactics. *Am J Med* 2004; 116(suppl 3A): 10S–16S.

14–16 June
Prague 2009
PERIOPERATIVE CARE
FOR THE GERIATRIC PATIENT

Abstracts wil be accepted for oral and/or poster presentations in the following areas:

- Perioperative outcomes in the geriatric population
- Postopperative analgesia for the aged
- Optimization of the treatment of comorbidities
- Prevention of infection and sepsis
- Optimizing perioperative patient management
- Early mobilization, ambulation and rehabilitation
- Regional anesthesia in the geriatric population
- ICU management for the elderly
- Day admission surgery for major operations in the elderly
- Issue of futility: when to withold or withdraw treatment
- Cognitive dysfunction and postoperative delirium economic considerations of geriatric care

The abstract submission deadline is Sunday, March 01, 2009. **Instructions are on website: www.geriatric09.cz.**
Abstract presenters must be registered for the meeting!

George Silvay, M.D., Ph.D.
Honorary President of the Congress
Professor, Department of Anesthesiology,
New York, The Mount Sinai School of Medicine New York
George.Silvay@mountsinai.org

Prof. Dr. Karel Cvachovec, CSc., MBA
Chair, Dept. of Anaesthesiology and CCM
Charles University 2nd School of Medicine Prague
Co-Chairman of the Organising Committee
Karel.Cvachovec@fnmotol.cz

Congress secretariat:
GUARANT International spol. s r.o.
Opletalova 22, 110 00 Prague, Czech Republic
Tel: +420 284 001 444, Fax: +420 284 001 448, E-mail: geriatric2009@guarant.cz

www.geriatric09.cz

▨ **HINARI PHASE I COUNTRIES**
Free access to more than 100 biomedical journals published by Wiley

□ **HINARI PHASE II COUNTRIES**
Nominal cost access to more than 100 biomedical journals published by Wiley

Developing World Access to Research

"perhaps the biggest step ever taken towards reducing the health information gap between rich and poor countries."

Former WHO Director-General Dr Gro Harlem Brundtland (July, 2001)

Wiley, together with the World Health Organisation and fellow Scientific, Technical and Medical publishers, is proud to play a role in furthering global dissemination of biomedical information.

Through the Health InterNetwork Access to Research Initiative (HINARI) www.healthinternetwork.org, free or nominal cost online access to more than 2,200 biomedical journals from 42 publishers is provided to more than 1,000 institutions in approximately 100 of the world's lowest income countries.

Wiley is one of the six founding publishers who led this initiative with the WHO. We offer more than 100 of our leading biomedical journals through the HINARI program; including many leading journals published with our society partners.

HINARI

WILEY
InterScience®
DISCOVER SOMETHING GREAT
www.interscience.wiley.com

Wiley – helping to make a world of difference.

5756

Giant
Resources...
Purebred
Content

Wiley InterScience® Reference Works

Sample Title List
Encyclopedia of Life Sciences
Encyclopedia of Statistical Sciences
Kirk-Othmer Encyclopedia of Chemical Technology
Sax's Dangerous Properties of Industrial Materials
Ullmann's Encyclopedia of Industrial Chemistry
Van Nostrand's Encyclopedia of Chemistry

Capabilities
Full text HTML and PDF (for some)
Reference linking
Citation download

Access
Single article or full title
One-time fee, subscription, or Pay-Per-View options
Unlimited concurrent usage

Today's sophisticated researcher demands fast and convenient online access to the giants of scientific reference. From *Kirk-Othmer* to the *Encyclopedia of Life Sciences*, **Wiley InterScience® Reference Works** deliver the biggest names in scientific research literature to your desktop, right when you need them. Encompassing subjects ranging from chemistry, life sciences, and medicine to psychology, engineering, and physics, these multi-volume encyclopedias, handbooks, and dictionaries are enhanced by powerful functionality that accelerates your path to discovery. With **Wiley InterScience**, you can be sure to ***Discover Something Great***.

WILEY InterScience®
DISCOVER SOMETHING GREAT

WILEY
1807-2007 KNOWLEDGE FOR GENERATIONS

www.interscience.wiley.com

LEVY LIBRARY, MT. SINAI SCHOOL O
3 4805 0808036 4

Fresh Content
Sliced Any Way You Like It

Wiley InterScience® OnlineBooks™

Content
Thousands of Wiley book titles across science, technology, medicine, business, and finance

Access
Single chapter or full title
One-time fee, subscription, or Pay-Per-View options

Capabilities
Search at the chapter level
Full text presented in PDF format
Citation download feature
COUNTER-compliant usage statistics

Satisfy your library's evolving research needs and budget with a customized blend of new content and licensing options at **Wiley InterScience® OnlineBooks™**. Pick and choose from an expanding roster of thousands of Wiley titles available online via **Wiley InterScience** to create a dynamic electronic book collection. Select from our new flexible licensing models – one-time fee, subscription, or Pay-Per-View. Add books as you need them from our distinguished portfolio of classic and cutting-edge titles. Whichever combination you select, enjoy unlimited concurrent access to content across the spectrum of science, medicine, technology, business, and finance.

WILEY
InterScience®
DISCOVER SOMETHING GREAT

WILEY
1807-2007 KNOWLEDGE FOR GENERATIONS

www.interscience.wiley.com

MOUNT SINAI
GUSTAVE L.
AND
JANET W.
LEVY
LIBRARY

MOUNT SINAI
GUSTAVE L.
AND
JANET W.
LEVY
LIBRARY

LEVY LIBRARY, MT. SINAI MEDICAL CENTER
3 4805 0562241 6